

Chapter 2

GAUSSIAN RANDOM VECTORS

2.1 Introduction

Gaussian random variables and Gaussian random vectors (vectors whose components are jointly Gaussian, as defined later) play a central role in detection and estimation. Part of the reason for this is that noise like quantities in many applications are reasonably modeled as Gaussian. Another, perhaps more important reason, is that Gaussian random variables turn out to be remarkably easy to work with (after an initial period of learning their peculiarities). Jointly Gaussian random variables are completely described by their means and covariances, which is part of the simplicity of working with them. When we find estimates or detection rules, and when we evaluate their performance, the answers then involve only those means and covariances.

A third reason why Gaussian random variables and vectors are so important is that we shall find, in many cases, that the performance measures we get for estimation and detection problems for the Gaussian case often bounds the performance for other random variables with the same means and covariances. For example, we will find that the minimum mean square estimator for Gaussian problems is the same, and has the same mean square performance, as the linear least squares estimator for other problems with the same mean and covariance. We will also find that this estimator is quite simple and is linear in the observations. Finally, we will find that the minimum mean square estimator for non-Gaussian problems always has a better performance than that for Gaussian problems of the same mean and covariance, but that the estimator is frequently much more complex. The point of this example is that non-Gaussian problems are often more easily and more deeply understood if we first understand the corresponding Gaussian problem.

In this chapter, we develop the most important properties of Gaussian random variables and vectors, namely the moment generating function, the moments, the joint densities, and the conditional probability densities. We also develop the properties of covariance matrices

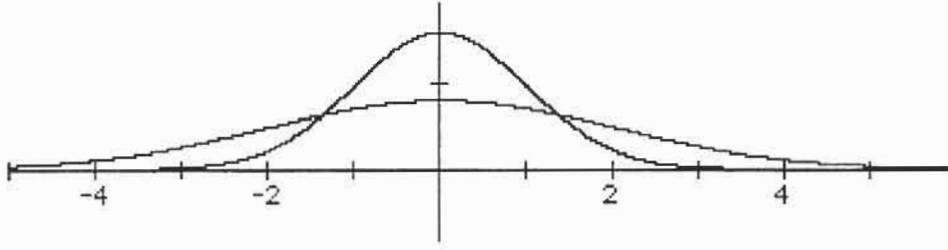


Figure 2.1: Graph of the density of a normalized Gaussian rv (the taller curve) and of a zero mean Gaussian rv with variance 4 (the flatter curve).

(which apply to both Gaussian and non-Gaussian random variables and vectors), and review a number of results about linear algebra that will be used in subsequent chapters.

2.2 Gaussian Random Variables

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

$$p_W(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) \quad (2.1)$$

Exercise 2.1 shows that $p_W(w)$ integrates to 1 (i.e., it is a probability density), and that W has mean 0 and variance 1. If we scale W by a positive constant σ to get the rv $Z = \sigma W$, then the density of Z at $z = \sigma w$ satisfies $p_Z(z)dz = p_W(w)dw$. Since $dz/dw = \sigma$, the density of Z is

$$p_Z(z) = \frac{1}{\sigma} p_W\left(\frac{z}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad (2.2)$$

Thus the density function for Z is scaled horizontally by the factor σ , and then scaled vertically by $1/\sigma$ (see Figure 2.1). 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 random variable for which $\Pr(Z=0) = 1$. For convenience in what follows, we use (2.2) 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 (2.2), for any $\sigma \geq 0$, is defined to be a *zero mean Gaussian rv*. The values $\Pr(|Z| < \sigma) = .682$, $\Pr(|Z| < 3\sigma) = .997$ give us a sense of how small the tails of the Gaussian distribution are.

If we shift Z to $U = Z + m$, then the density shifts so as to be centered at m , the mean becomes m , and the density satisfies $p_U(u) = p_Z(u - m)$, so that

$$p_U(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u - m)^2}{2\sigma^2}\right) \quad (2.3)$$

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

The added generality of a mean often obscures formulas; we will usually work with zero mean rv's and random vectors (rv's) and insert the means later. That is, any random variable can be regarded as a constant (the mean) plus a zero mean random variable (called the fluctuation). When necessary for notation, we denote $U = m_U + \tilde{U}$, where \tilde{U} is the fluctuation of U around its mean m_U , and work with \tilde{U} .

The *moment generating function* (MGF) of an arbitrary rv Z is defined to be $g_Z(s) = E[\exp(sZ)]$. We usually take s to be a real variable, but it can also be regarded as complex, with real part α and imaginary part $j\omega$. The two sided Laplace transform of the density of Z is equal to $g_Z(-s)$. Similarly, the Fourier transform of the density of Z , as a function of ω , is $g_Z(-j\omega)$ and the characteristic function of Z is $g_Z(j\omega)$. The characteristic function and Fourier transform have the advantage that, for real ω , they exist for all random variables, whereas the MGF and Laplace transform (for real $s \neq 0$) exist only if the tails of the distribution approach 0 at least exponentially. For the rv's of interest here, the MGF exists for at least a region of real s around 0, and thus, if one calculates one of these transforms, say the MGF, the others follow simply by substituting $-s$, $j\omega$, or $-j\omega$ for s .

For the Gaussian rv $Z \sim \mathcal{N}(0, \sigma^2)$, $g_Z(s)$ can be calculated as follows:

$$\begin{aligned} g_Z(s) &= E[\exp(sZ)] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(sz) \exp\left[-\frac{z^2}{2\sigma^2}\right] dz \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[\frac{-z^2 + 2\sigma^2 sz - s^2\sigma^4}{2\sigma^2} + \frac{s^2\sigma^2}{2}\right] dz \end{aligned} \quad (2.4)$$

$$= \exp\left[\frac{s^2\sigma^2}{2}\right] \left\{ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(z - s\sigma)^2}{2\sigma^2}\right] dz \right\} \quad (2.5)$$

$$= \exp\left[\frac{s^2\sigma^2}{2}\right] \quad (2.6)$$

We completed the square in the exponent in (2.4) and then recognized, in (2.5), that the term in braces is the integral of a probability density and thus equal to 1. If the MGF of Z exists for a region of real s around 0 (as it does for Gaussian rv's), it can be used to calculate all the moments of Z . As shown in Exercise 2.2, $E[Z^{2k}]$, for $Z \sim \mathcal{N}(0, \sigma^2)$, is given, for all integers $k \geq 1$, by

$$E[Z^{2k}] = \frac{(2k)!\sigma^{2k}}{k!2^k} = (2k-1)(2k-3)(2k-5)\dots(3)(1)\sigma^{2k} \quad (2.7)$$

Thus, $E[Z^4] = 3\sigma^4$, $E[Z^6] = 15\sigma^6$, etc. Since z^{2k+1} is an odd function of z and the Gaussian density is an even function, the odd moments of Z are all zero. For an arbitrary Gaussian rv $U \sim \mathcal{N}(m, \sigma^2)$, we have $U = m + \tilde{U}$, where $\tilde{U} \sim \mathcal{N}(0, \sigma^2)$. Thus $g_U(s) = E[\exp(s(m + \tilde{U}))] = e^{sm} E[s\tilde{U}] = e^{sm} \exp(s^2\sigma^2/2)$. It turns out that the distribution function of a random variable is uniquely determined¹ by its moment generating function

¹More precisely, the distribution function is uniquely determined except on a set of points of measure zero.

(if the MGF exists in an interval around 0), so we see that a random variable U is Gaussian with mean m and variance σ^2 if and only if (iff) its moment generating function is

$$g_Z(s) = \exp\left(sm + \frac{s^2\sigma^2}{2}\right) \quad (2.8)$$

2.3 Gaussian Random Vector and MGF's

An n by m *matrix* A is an array of nm elements arranged in n rows and m columns; a_{ij} denotes the j^{th} element in the i^{th} row. Unless specified to the contrary, the elements will be real numbers. The *transpose* A^T of an n by m matrix A is an m by n matrix B with $b_{ji} = a_{ij}$ for all i, j . A matrix is *square* if $n = m$ and a square matrix A is *symmetric* if $A = A^T$. If A and B are each n by m matrices, $A + B$ is an n by m matrix C with $c_{ij} = a_{ij} + b_{ij}$ for all i, j . If A is n by m and B is m by r , the matrix AB is an n by r matrix C with elements $c_{ik} = \sum_j a_{ij}b_{jk}$. 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 \vec{a} as $(a_1, \dots, a_n)^T$. The reader is expected to be familiar with vector and matrix manipulations.

An n -dimensional random vector (an n -rv) is a mapping from the sample space into the space \mathbb{R}^n of n -dimensional real vectors. We could view an n -rv simply as n individual random variables, but vector notation allows us to state results much more compactly for vectors than for the set of individual rv's. Sampled time stochastic processes can be viewed simply as random vectors (although the dimension is often infinite), and continuous time stochastic processes are usually best studied by various expansions that transform them to rv's (although again often of infinite dimension). Thus, a thorough understanding of rv's is essential to everything that follows in this subject.

The *probability density*, $p_{\vec{Z}}(\vec{z})$, of an n -rv $\vec{Z} = (Z_1, Z_2, \dots, Z_n)^T$ is simply the joint probability density of the components Z_1, \dots, Z_n . The *mean* of \vec{Z} , denoted $\vec{m}_{\vec{Z}}$ or $E[\vec{Z}]$, is the real vector $(m_{Z_1}, m_{Z_2}, \dots, m_{Z_n})^T$ where $m_{Z_i} = E[Z_i]$ for $1 \leq i \leq n$. Similarly the covariance matrix of \vec{Z} , denoted $K_{\vec{Z}}$, is defined as $E[(\vec{Z} - \vec{m}_{\vec{Z}})(\vec{Z} - \vec{m}_{\vec{Z}})^T]$. This is an n by n symmetric matrix whose element in the i th row, j th column, is the covariance of Z_i and Z_j , i.e., $E[(Z_i - m_i)(Z_j - m_j)]$. Finally, the *moment generating function* (MGF) of an n -rv \vec{Z} is defined as $g_{\vec{Z}}(\vec{s}) = E[\exp(\vec{s}^T \vec{Z})]$ where $\vec{s} = (s_1, \dots, s_n)^T$ is an n -dimensional vector. The components of \vec{s} could be taken to be complex, but we usually view them as real. If the components of a rv are independent and identically distributed (IID), we call the vector an IID rv.

Example 2.1 An example that will become very familiar is that of an IID n -rv \vec{W} where the components W_i , $1 \leq i \leq n$, are each normalized Gaussian, $W_i \sim \mathcal{N}(0, 1)$. By taking the product of n densities given by (2.1), the density of $\vec{W} = (W_1, W_2, \dots, W_n)^T$ is

$$p_{\vec{W}}(\vec{w}) = \frac{1}{(2\pi)^{n/2}} \exp\left(\frac{-w_1^2 - w_2^2 - \dots - w_n^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(\frac{-\vec{w}^T \vec{w}}{2}\right) \quad (2.9)$$

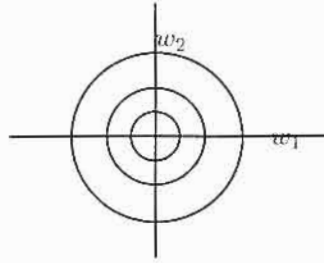


Figure 2.2: Contours of equal probability density for two normalized IID Gaussian rv's.

The joint density of \vec{W} at a sample value \vec{w} depends only on the squared distance $\vec{w}^T \vec{w}$ of the sample value from the origin. That is, $p_{\vec{W}}(\vec{w})$ is spherically symmetric around the origin, and points of equal probability density lie on concentric spheres around the origin (see Figure 2.2).

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

$$\begin{aligned} g_{\vec{W}}(\vec{s}) &= E[\exp \vec{s}^T \vec{W}] = E[\exp(s_1 W_1 + \cdots + s_n W_n)] = E\left[\prod_i \exp(s_i W_i)\right] \\ &= \prod_i E[\exp(s_i W_i)] = \prod_i \exp\left(\frac{s_i^2}{2}\right) = \exp\left[\frac{\vec{s}^T \vec{s}}{2}\right], \end{aligned} \quad (2.10)$$

where we have used, first, the fact that the independence of $\{W_1, \dots, W_n\}$ guarantees the independence of $\{\exp(s_1 W_1), \dots, \exp(s_n W_n)\}$, next, the fact that the expected value of a product of independent rv's is equal to the product of the expected values, and, finally, the fact that (2.6) gives the MGF of each W_i .

We now go on to define the general class of Gaussian rv's.

Definition: $\{Z_1, Z_2, \dots, Z_n\}$ is a set of jointly Gaussian random variables, and $\vec{Z} = (Z_1, \dots, Z_n)^T$ is a *Gaussian* rv, if, for all real vectors $\vec{s} = (s_1, \dots, s_n)^T$, the linear combination $\vec{s}^T \vec{Z} = s_1 Z_1 + s_2 Z_2 + \cdots + s_n Z_n$ is a Gaussian random variable.

The intuitive idea here is that Gaussian rv's arise in practice because of the addition of large numbers of small essentially independent rv's (the central limit theorem indicates that such a sum can be approximated by a Gaussian rv). For example, when a broad band noise waveform is passed through a narrow band linear filter, the output at any given time is usually well approximated as a Gaussian rv. A linear combination of outputs at different times is also a sum of the same set of small, essentially independent, underlying rv's, and that sum can again be approximated as Gaussian. Thus we would expect a set of outputs at different times to be a jointly Gaussian set according to the above definition.

Note that if $\{Z_1, \dots, Z_n\}$ is jointly Gaussian, then for each i , $1 \leq i \leq n$, Z_i is a Gaussian random variable (as follows by choosing $s_i = 1$ and $s_j = 0$ for $j \neq i$ in the definition above). In the same way, each subset of $\{Z_1, \dots, Z_n\}$ is also jointly Gaussian. However, if Z_1, Z_2, \dots, Z_n are each Gaussian rv's, it does not necessarily follow that the set $\{Z_1, \dots, Z_n\}$

is jointly Gaussian. As a simple example, let $Z_1 \sim \mathcal{N}(0, 1)$, let X be $+1$ or -1 , each with probability $1/2$, and let $Z_2 = Z_1 X_1$. Then $Z_2 \sim \mathcal{N}(0, 1)$. The joint probability density, $p_{Z_1 Z_2}(z_1, z_2)$ is then 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 half². Exercise 2.3 gives another example. The distinction between individually Gaussian and jointly Gaussian rv's is much more than mathematical nit picking, since the remarkable properties of Gaussian random vectors follow largely from the jointly Gaussian property rather than merely the property of being individually Gaussian.

We now find the moment generating function (MGF) of the rv $\vec{Z} = (Z_1, \dots, Z_n)^T$ under the assumption that \vec{Z} is a zero mean Gaussian rv. The MGF, by definition, is

$$g_{\vec{Z}}(\vec{s}) = E[\exp(\vec{s}^T \vec{Z})] = E\left[\exp\left(\sum_i s_i Z_i\right)\right]$$

For any given \vec{s} , let $X = \vec{s}^T \vec{Z}$. Since \vec{Z} is a Gaussian rv (i.e., (Z_1, \dots, Z_n) is jointly Gaussian), X is also Gaussian, and since \vec{Z} is zero mean (i.e., all components are zero mean), X is zero mean. From (2.6), $g_X(r) = \exp[r^2 \sigma_X^2 / 2]$, so that $g_X(1) = \exp[\sigma_X^2 / 2]$. For $X = \vec{s}^T \vec{Z}$ then,

$$g_{\vec{Z}}(\vec{s}) = E[\exp(\vec{s}^T \vec{Z})] = E[\exp(X)] = g_X(1) = \exp(\sigma_X^2 / 2)$$

Finally $\sigma_X^2 = E[X^2] = E[\vec{s}^T \vec{Z} \vec{Z}^T \vec{s}] = \vec{s}^T K_{\vec{Z}} \vec{s}$, where $K_{\vec{Z}}$ is the covariance matrix of \vec{Z} . Thus, the moment generating function of an arbitrary zero mean Gaussian rv \vec{Z} is

$$g_{\vec{Z}}(\vec{s}) = \exp\left[\frac{\vec{s}^T K_{\vec{Z}} \vec{s}}{2}\right] \quad (2.11)$$

Next let $\vec{U} = (U_1, \dots, U_n)^T$ be an arbitrary Gaussian rv and, for $1 \leq i \leq n$, let $m_i = E[U_i]$ and let $Z_i = U_i - m_i$ be the fluctuation of U_i . Letting $\vec{m} = (m_1, \dots, m_n)^T$ and $\vec{Z} = (Z_1, \dots, Z_n)$, we have $\vec{U} = \vec{m} + \vec{Z}$. Thus the MGF of \vec{U} is given by $g_{\vec{U}}(\vec{s}) = E[\exp(\vec{s}^T \vec{m} + \vec{s}^T \vec{Z})] = \exp(\vec{s}^T \vec{m}) g_{\vec{Z}}(\vec{s})$. Using (2.11) and recognizing that \vec{Z} and \vec{U} have the same covariance function,

$$g_{\vec{U}}(\vec{s}) = \exp\left(\vec{s}^T \vec{m} + \frac{\vec{s}^T K_{\vec{U}} \vec{s}}{2}\right) \quad (2.12)$$

Note that the MGF of a Gaussian rv is completely specified by its mean and covariance. Because of this, we denote a Gaussian rv \vec{U} of mean \vec{m} and covariance $K_{\vec{U}}$ as $\vec{U} \sim \mathcal{N}(\vec{m}, K_{\vec{U}})$. Note also that if a rv \vec{U} has the MGF in (2.12), then, as shown in Exercise 2.5, each linear combination of the components of \vec{U} is a Gaussian rv. Thus we have the theorem,

Theorem 2.1 A rv \vec{U} with mean \vec{m} and covariance $K_{\vec{U}}$ is a Gaussian rv iff $g_{\vec{U}}(\vec{s})$ is given by (2.12).

²One frequently hears the erroneous statement that uncorrelated Gaussian rv's are independent. This is false as the above example shows. The correct statement, as we see later, is that uncorrelated *jointly* Gaussian rv's are independent.

Note that the MGF's in (2.6), (2.8), and (2.11) are special cases of (2.12). Also the MGF of the IID rv \vec{W} in (2.10) agrees with (2.12) since the mean of \vec{W} is zero and the covariance is the identity matrix I . Thus \vec{W} , not surprisingly, is a Gaussian rv, i.e., $\vec{W} \sim \mathcal{N}(0, I)$.

2.4 Joint Probability Densities for Gaussian Random Vectors

In this section, we start with an example of a zero mean Gaussian rv that is defined as a linear transformation of the IID normalized Gaussian rv \vec{W} of example 2.1. We show later that this example in fact covers all zero mean Gaussian rv's.

Example 2.2 Let A be an n by n real matrix, let \vec{W} be an IID normalized Gaussian n -rv, and let the n -rv \vec{Z} be given by $\vec{Z} = A\vec{W}$. The covariance matrix of \vec{Z} is

$$K_{\vec{Z}} = E[\vec{Z}\vec{Z}^T] = E[A\vec{W}\vec{W}^T A^T] = AA^T \quad (2.13)$$

since $E[\vec{W}\vec{W}^T]$ is the identity matrix, I_n . We can easily find the MGF of \vec{Z} since

$$\begin{aligned} g_{\vec{Z}}(\vec{s}) &= E[\exp(\vec{s}^T \vec{Z})] = E[\exp(\vec{s}^T A\vec{W})] = E[\exp\{(A^T \vec{s})^T \vec{W}\}] = g_{\vec{W}}(A^T \vec{s}) \\ &= \exp\left[\frac{\vec{s}^T A A^T \vec{s}}{2}\right] = \exp\left[\frac{\vec{s}^T K_{\vec{Z}} \vec{s}}{2}\right] \end{aligned} \quad (2.14)$$

Comparing (2.14) with (2.11), we see that \vec{Z} is a zero mean Gaussian rv. We shall see shortly that any zero mean Gaussian rv can be represented in the form $\vec{Z} = A\vec{W}$ for some real n by n matrix A and IID normalized Gaussian rv \vec{W} .

We next find the joint probability density of $\vec{Z} = A\vec{W}$. First consider the corresponding transformation of real valued vectors, $\vec{z} = A\vec{w}$. Let \vec{e}_i be the i th unit vector (i.e., the vector whose i th component is 1 and whose other components are 0). Then $A\vec{e}_i = \vec{a}_i$, where \vec{a}_i is the i th column of A . Thus, $\vec{z} = A\vec{w}$ transforms the unit vectors \vec{e}_i into the columns \vec{a}_i of A . For $n = 2$, Figure 2.3 shows how this transformation carries the unit square with the corners $\vec{0}, \vec{e}_1, \vec{e}_2$, and $(\vec{e}_1 + \vec{e}_2)$ into the parallelogram with corners $\vec{0}, \vec{a}_1, \vec{a}_2$, and $(\vec{a}_1 + \vec{a}_2)$.

For an arbitrary number of dimensions, the unit cube in the \vec{w} space is the set of points \vec{w} such that $0 \leq w_i \leq 1$ for $i = 1, \dots, n$. 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 $\vec{e}_{i_1} + \vec{e}_{i_2} + \dots + \vec{e}_{i_k}$. The transformation $A\vec{w}$ carries the unit cube into a parallelepiped, where each corner of the cube, $\vec{e}_{i_1} + \vec{e}_{i_2} + \dots + \vec{e}_{i_k}$, is carried into a corresponding corner $\vec{a}_{i_1} + \vec{a}_{i_2} + \dots + \vec{a}_{i_k}$ of the parallelepiped. One of the most interesting and geometrically meaningful properties of the determinant, $\det(A)$, of a square matrix A is that the *magnitude* of that determinant, $|\det(A)|$, is equal to the volume of that parallelepiped (see Strang, Linear Algebra, sec. 4.4). If $\det(A) = 0$, i.e., if A is singular, then the n -dimensional unit cube in the w space

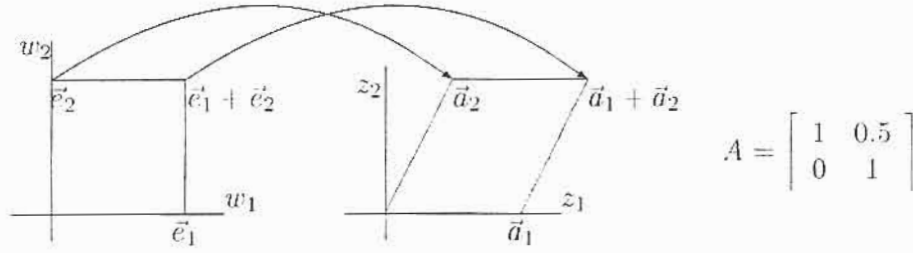


Figure 2.3: Unit square transformed to a parallelogram.

is transformed into a smaller dimensional parallelepiped whose volume (as a region of n -dimensional space) is 0. We assume in what follows that $\det(A) \neq 0$ so that the inverse of A exists.

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

$$p_{\vec{Z}}(\vec{z})|d\vec{z}| = p_{\vec{W}}(\vec{w})|d\vec{w}| \quad (2.15)$$

where $|d\vec{w}|$ is the volume of an incremental cube with dimension $\delta = dw_i$ on each side, and $|d\vec{z}|$ is the volume of that incremental cube transformed by A . Thus $|d\vec{w}| = \delta^n$ and $|d\vec{z}| = \delta^n |\det(A)|$ so that $|d\vec{z}|/|d\vec{w}| = |\det(A)|$. Using this in (2.15), and using (2.9) for $p_{\vec{W}}(\vec{w}) = p_{\vec{W}}(A^{-1}\vec{z})$, we see that the density of a jointly Gaussian vector $\vec{Z} = A\vec{W}$ is

$$p_{\vec{Z}}(\vec{z}) = \frac{\exp\left[-\frac{1}{2}\vec{z}^T(A^{-1})^T A^{-1}\vec{z}\right]}{(2\pi)^{n/2}|\det(A)|} \quad (2.16)$$

From (2.13), we have $K_{\vec{Z}} = AA^T$, so $K_{\vec{Z}}^{-1} = (A^{-1})^T A^{-1}$ and $\det(K_{\vec{Z}}) = \det(A)\det(A^T) = [\det(A)]^2 > 0$. Thus (2.16) becomes

$$p_{\vec{Z}}(\vec{z}) = \frac{\exp\left[-\frac{1}{2}\vec{z}^T K_{\vec{Z}}^{-1}\vec{z}\right]}{(2\pi)^{n/2}\sqrt{\det(K_{\vec{Z}})}} \quad (2.17)$$

Eq. (2.17) doesn't have any meaning when A , and thus $K_{\vec{Z}}$, is singular, since $K_{\vec{Z}}^{-1}$ does not then exist. In the case where A is singular, $A\vec{w}$ maps the set of n -dimensional vectors \vec{w} into a subspace of dimension less than n , and $p_{\vec{Z}}(\vec{z})$ is 0 outside of that subspace and is impulsive inside. What this means is that some components of the random vector \vec{Z} can be expressed as linear combinations of other components. In this case, one can avoid very messy notation by simply defining the components that are linearly independent as forming a random vector \vec{Z}' , and representing the other components as linear combinations of the components of \vec{Z}' . With this stratagem, we can always take the covariance of any such reduced rv to be nonsingular.

Next consider $\vec{U} = \vec{m} + A\vec{W}$. Define $\vec{U} = A\vec{W}$ as the fluctuation in \vec{U} , and note that (2.17) can be used for the density of \vec{U} . Since $E[\vec{U}] = \vec{0}$, we see that $E[\vec{U}] = \vec{m}$. Assuming

$\det(A) \neq 0$, we can immediately write down the density for \vec{U} as a translation of that for \vec{U} ,

$$p_{\vec{U}}(\vec{u}) = \frac{\exp\left[-\frac{1}{2}(\vec{u} - \vec{m})^T K_{\vec{U}}^{-1}(\vec{u} - \vec{m})\right]}{(2\pi)^{n/2} \sqrt{\det(K_{\vec{U}})}} \quad (2.18)$$

where $K_{\vec{U}} = E[\vec{U} \vec{U}^T]$ is the covariance matrix of both \vec{U} and \vec{U} . We soon show that this is the general form of density for a Gaussian random vector $\vec{U} \sim \mathcal{N}(\vec{m}, K_{\vec{U}})$ if $\det(K_{\vec{U}}) \neq 0$. In fact, we have already shown that for any Gaussian rv $\vec{U} \sim \mathcal{N}(\vec{m}, K_{\vec{U}})$, the MGF satisfies (2.12), and it turns out that this uniquely specifies the distribution of \vec{U} . Thus, if there is a matrix A with $\det(A) \neq 0$ such that $K_{\vec{U}} = AA^T$, then $\vec{U} \sim \mathcal{N}(\vec{m}, K_{\vec{U}})$ can be expressed as $\vec{m} + A\vec{W}$, and (2.18) must be the density of \vec{U} . We shall soon see that any covariance matrix $K_{\vec{U}}$ can be expressed as AA^T for some square matrix A . In most of what follows, we restrict our attention to zero mean Gaussian rv's, since, as we have shown, it is usually simpler to deal with the fluctuation \vec{U} of \vec{U} , and then include the mean later.

For the 2-dimensional zero mean case, let $E[Z_1^2] = \sigma_1^2$, $E[Z_2^2] = \sigma_2^2$ and $E[Z_1 Z_2] = k_{12}$. Define the *normalized covariance*, ρ , as $k_{12}/(\sigma_1 \sigma_2)$. Then $\det(K_{\vec{Z}}) = \sigma_1^2 \sigma_2^2 - k_{12}^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$. For A to be non-singular, we need $\det(K_{\vec{Z}}) = [\det(A)]^2 > 0$, so we need $|\rho| < 1$. We then have

$$\begin{aligned} K_{\vec{Z}}^{-1} &= \frac{1}{\sigma_1^2 \sigma_2^2 - k_{12}^2} \begin{bmatrix} \sigma_2^2 & -k_{12} \\ -k_{12} & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \rho^2} \begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1 \sigma_2) \\ -\rho/(\sigma_1 \sigma_2) & 1/\sigma_2^2 \end{bmatrix} \\ p_{\vec{Z}}(\vec{z}) &= \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2 - k_{12}^2}} \exp\left(\frac{-z_1^2 \sigma_2^2 + 2z_1 z_2 k_{12} - z_2^2 \sigma_1^2}{2(\sigma_1^2 \sigma_2^2 - k_{12}^2)}\right) \\ &= \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left(\frac{-(z_1/\sigma_1)^2 + 2\rho(z_1/\sigma_1)(z_2/\sigma_2) - (z_2/\sigma_2)^2}{2(1 - \rho^2)}\right) \end{aligned} \quad (2.19)$$

This is why we use vector notation! There are two lessons in this. First, hand calculation is frequently messy for Gaussian rv's, and second, the vector equations are much simpler, so we must learn to reason directly from the vector equations and use standard computer programs to do the calculations.

2.5 Properties of Covariance Matrices

In this section, we summarize some properties of covariance matrices that will be used frequently in what follows. A matrix K is a *covariance matrix* if a zero mean rv \vec{Z} exists such that $K = E[\vec{Z} \vec{Z}^T]$. It is important to realize that the properties developed here apply to non-Gaussian as well as Gaussian rv's. An n by n matrix K is *positive semi-definite* if it is symmetric and if, for all real n -vectors \vec{b} , $\vec{b}^T K \vec{b} \geq 0$. It is *positive definite* if, in addition, $\vec{b}^T K \vec{b} > 0$ for all non-zero \vec{b} . Our objective in this section is to list the relationships between

these types of matrices, and to state some other frequently useful properties of covariance matrices.

1) Every covariance matrix K is positive semi-definite. To see this, let \vec{Z} be a zero mean n -rv such that $K = E[\vec{Z}\vec{Z}^T]$. K is symmetric since $E[Z_i Z_j] = E[Z_j Z_i]$ for all i, j . Let \vec{b} be an arbitrary real n -vector, and let $X = \vec{b}^T \vec{Z}$. Then $0 \leq E[X^2] = E[\vec{b}^T \vec{Z} \vec{Z}^T \vec{b}] = \vec{b}^T K \vec{b}$.

2) A covariance matrix K is positive definite iff $\det(K) \neq 0$. To see this, define \vec{Z} as above and note that if $\vec{b}^T K \vec{b} = 0$ for some $\vec{b} \neq 0$, then $X = \vec{b}^T \vec{Z}$ has zero variance, and therefore is zero with probability 1. Thus $E[X \vec{Z}^T] = 0$, so $\vec{b}^T E[\vec{Z} \vec{Z}^T] = 0$. Since $\vec{b} \neq 0$ and $\vec{b}^T K = 0$, we must have $\det(K) = 0$. Conversely, if $\det(K) = 0$, there is some \vec{b} such that $K \vec{b} = 0$, so $\vec{b}^T K \vec{b}$ is also 0.

3) A complex number λ is an *eigenvalue* of a matrix K if $K \vec{q} = \lambda \vec{q}$ for some non-zero vector \vec{q} ; the corresponding \vec{q} is called an *eigenvector*. The following results about the eigenvalues and eigenvectors of positive definite (semi-definite) matrices K are standard linear algebra results (see for example, Strang, section 5.5)³

All eigenvalues of K are positive (non-negative). All the eigenvectors can be taken to be real. All eigenvectors of different eigenvalues are orthogonal, and if an eigenvalue has multiplicity j , then it has j orthogonal eigenvectors. Altogether, n orthogonal eigenvectors can be chosen, and they can be scaled to be orthonormal.

4) If K is positive semi-definite, there is an orthonormal matrix Q whose columns, $\vec{q}_1, \dots, \vec{q}_n$ are the orthonormal eigenvectors above. Q satisfies $KQ = Q\Lambda$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix whose i th element, λ_i , is the eigenvalue of K corresponding to the eigenvector \vec{q}_i . This is simply the vector version of the eigenvector/eigenvalue relationship in property 3. Q also satisfies $Q^T Q = I$ where I is the identity matrix. This follows since $\vec{q}_i^T \vec{q}_j = \delta_{ij}$. We then also have $Q^{-1} = Q^T$.

5) If K is positive semi-definite, and if Q and Λ are the matrices above, then $K = Q\Lambda Q^T$. If K is positive definite, then also $K^{-1} = Q\Lambda^{-1}Q^T$. The first equality follows from property 4, and the second follows by multiplying the expression for K by that for K^{-1} , which exists because all the eigenvalues are positive.

6) For a symmetric n by n matrix K , $\det K = \prod_{i=1}^n \lambda_i$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of K repeated according to their multiplicity. Thus if K is positive definite, $\det K > 0$ and if K is positive semi-definite, $\det K \geq 0$.

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

$$R = Q\Lambda^{1/2}Q^T \text{ where } \Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}) \quad (2.20)$$

³Strang also shows that K is positive definite iff all the upper left square submatrices have positive determinants. This is often called Sylvester's test. This test does not extend to positive semi-definite matrices (for example, the matrix $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ is not positive semi-definite even though the upper left square submatrices have non-negative determinants). K is positive semi-definite iff all the principal submatrices (i.e., the submatrices resulting from dropping a subset of rows and the corresponding subset of columns) have non-negative determinants.

8) If K is positive semi-definite, then K is a covariance matrix. In particular, K is the covariance matrix of $\vec{Z} = R\vec{W}$ where R is the square root matrix in (2.20) and \vec{W} is an IID normalized Gaussian r.v.

9) For any real n by n matrix A , the matrix $K = AA^T$ is a covariance matrix. In particular, K is the covariance matrix of $\vec{Z} = A\vec{W}$.

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

Theorem 2.2 A real n by n matrix K is a covariance matrix iff it is positive semi-definite. Also it is a covariance matrix iff $K = AA^T$ for some real n by n matrix A . One choice for A is the square root matrix R in (2.20).

We now see, as summarized in the following corollary, that example 2.2 above is completely general.

COROLLARY 1: For any covariance matrix K , a zero mean Gaussian r.v. $\vec{Z} \sim \mathcal{N}(0, K)$ exists and \vec{Z} can be viewed as $A\vec{W}$, where $AA^T = K$ and $\vec{W} \sim \mathcal{N}(0, I_n)$. The MGF of \vec{Z} is given by (2.11), and, if $\det(K) \neq 0$, the probability density is given by (2.17). The density for an arbitrary Gaussian r.v. $\vec{U} \sim (\vec{m}, K)$, with $\det(K) \neq 0$, is given by (2.18).

2.6 Geometry and Principal Axes for Gaussian Densities

Let \vec{Z} be an arbitrary zero mean Gaussian n -r.v. with non-singular covariance K . For any given $c > 0$, the points \vec{z} for which $\vec{z}^T K \vec{z} = c$ form a contour of equal probability density for \vec{Z} , as seen by (2.17). This is a quadratic equation in \vec{z} and is the equation for an ellipsoid centered on the origin. The principal axes of this ellipsoid are given by the eigenvectors of K . In order to understand this, let Q be an orthonormal matrix whose columns, $\vec{q}_1, \dots, \vec{q}_n$ are the eigenvectors of K and let Λ be the corresponding diagonal matrix of eigenvalues (as in property 4 above). Consider the transformation $\vec{V} = Q^T \vec{Z}$. The covariance matrix of \vec{V} is then

$$K_{\vec{V}} = E[\vec{V}\vec{V}^T] = E[Q^T \vec{Z} \vec{Z}^T Q] = Q^T K Q = \Lambda, \quad (2.21)$$

where the last step follows from $K = Q\Lambda Q^T$ and $Q^T = Q^{-1}$. Since \vec{Z} is a zero mean Gaussian r.v., \vec{V} is also, so it has the joint probability density

$$p_{\vec{V}}(\vec{v}) = \frac{\exp[-\vec{v}^T K_{\vec{V}}^{-1} \vec{v}/2]}{(2\pi)^{n/2} [\det(K_{\vec{V}})]^{1/2}} = \frac{\exp[-\sum_i v_i^2/(2\lambda_i)]}{(2\pi)^{n/2} [\prod_i \sqrt{\lambda_i}]} = \prod_{i=1}^n \frac{\exp[-v_i^2/(2\lambda_i)]}{\sqrt{2\pi\lambda_i}} \quad (2.22)$$

The transformation $\vec{V} = Q^T \vec{Z}$ carries each sample value \vec{z} for the r.v. \vec{Z} into the sample value $\vec{v} = Q^T \vec{z}$ of \vec{V} . Visualize this transformation as a change of basis, i.e., \vec{z} represents some point \vec{x} of n -dimensional real space in terms of a given co-ordinate system and \vec{v}

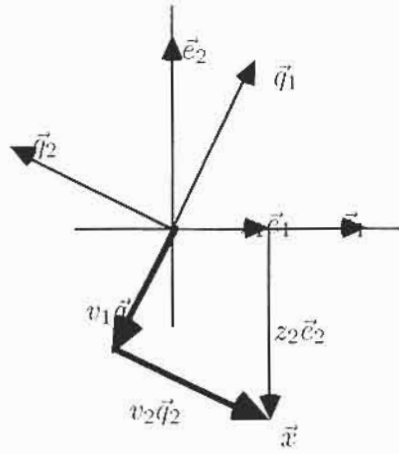


Figure 2.4: Representation of a point in two different co-ordinate systems. Point \vec{x} is represented by $(z_1, z_2)^T$ in the \vec{e}_1, \vec{e}_2 system and by $(v_1, v_2)^T$ in the \vec{q}_1, \vec{q}_2 system.

represents the same \vec{x} in a different co-ordinate system (see Figure 2.4). In particular, let $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ be the basis vectors in co-ordinate system 1, and let $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n$ be the basis vectors in co-ordinate system 2. Then

$$\vec{x} = \sum_{i=1}^n z_i \vec{e}_i = \sum_{i=1}^n v_i \vec{q}_i \quad (2.23)$$

In co-ordinate system 1, \vec{e}_i is simply the i th unit vector, containing 1 in position i and 0 elsewhere. In this co-ordinate system, \vec{x} is represented by the n -tuple $(z_1, \dots, z_n)^T$. Also, in co-ordinate system 1, \vec{q}_i is the i th normalized eigenvector of the matrix K . Since the orthonormal matrix Q has the columns $\vec{q}_1, \dots, \vec{q}_n$, the right hand side of (2.23) can be written, in co-ordinate system 1, as $Q\vec{v}$, where \vec{v} is the n -tuple $(v_1, \dots, v_n)^T$. Thus, in co-ordinate system 1, $\vec{z} = Q\vec{v}$, so $\vec{v} = Q^{-1}\vec{z} = Q^T\vec{z}$, verifying that $\vec{v} = Q^T\vec{z}$ actually corresponds to the indicated change of basis. Note that $\vec{q}_1, \dots, \vec{q}_n$ are orthonormal in both co-ordinate systems, and $\vec{e}_1, \dots, \vec{e}_n$ are also orthonormal in both systems. This means that the transformation can be viewed as a rotation, i.e., distances are unchanged (see Exercise 2.7).

For any given real number $c > 0$, the set of vectors \vec{z} for which $\vec{z}^T K^{-1} \vec{z} = c$ form a contour of equal probability density for \vec{Z} (in co-ordinate system 1). The corresponding contour in co-ordinate system 2, with $\vec{z} = Q\vec{v}$ is given by $\vec{v}^T Q^T K^{-1} Q \vec{v} = c$. This can be rewritten as $\vec{v}^T \Lambda^{-1} \vec{v} = c$, or $\sum_i v_i^2 / \lambda_i = c$ (as can also be seen from (2.22)). This is the equation of an ellipsoid, but the principal axes of the ellipse are now aligned with co-ordinate system 2, using basis vectors $\vec{q}_1, \dots, \vec{q}_n$ (see Figure 2.5). The distances from the origin to the ellipsoid in these principal axis directions are $\sqrt{c\lambda_i}$, $1 \leq i \leq n$.

Figure 2.5: Lines of equal probability density form ellipses. The principal axes of these ellipses are the eigenvectors of K . The intersection of an ellipse with each axis is proportional to the square root of the corresponding eigenvalue.

2.7 Conditional Probabilities

Next consider the conditional probability $p_{X|Y}(x|y)$ for two jointly Gaussian zero mean rv's X and Y with a non-singular covariance matrix. From (2.19),

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

where $\rho = E[XY]/(\sigma_X\sigma_Y)$. Since

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

and $Y \sim \mathcal{N}(0, \sigma_Y^2)$, we have

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

This simplifies to

$$p_{X|Y}(x|y) = \frac{1}{\sigma_X\sqrt{2\pi(1-\rho^2)}} \exp \left[\frac{-|x - \rho(\sigma_X/\sigma_Y)y|^2}{2\sigma_X^2(1-\rho^2)} \right] \quad (2.24)$$

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 $Y = y$, and in that restricted sample space, X is $\mathcal{N}(\rho(\sigma_X/\sigma_Y)y, \sigma_X^2(1-\rho^2))$. This means that the fluctuation of X , in this restricted space, has the same density for all y . When we study estimation, we shall find that the facts that, first, the conditional mean is linear in y , and, second, the fluctuation is independent of y , are crucially important. These simplifications will lead to many important properties and insights in what follows. We now go on to show that the 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 \vec{U} be an $n + m$ dimensional Gaussian rv. View \vec{U} as a pair of rv's \vec{X}, \vec{Y} of dimensions m and n respectively. That is, $\vec{U}^T = (U_1, U_2, \dots, U_{n+m}) = (X_1, X_2, \dots, X_m, Y_1, \dots, Y_n) = (\vec{X}^T, \vec{Y}^T)$. \vec{X} and \vec{Y} are called *jointly Gaussian* rv's. If the covariance matrix of \vec{U} is non-singular, we say that \vec{X} and \vec{Y} are *jointly non-singular*. In what follows, we assume that \vec{X} and \vec{Y} are jointly Gaussian, jointly non-singular, and zero-mean⁴. The covariance matrix $K_{\vec{U}}$ of \vec{U} can be partitioned into m rows on top and n rows on bottom, and then further partitioned into m and n columns, yielding:

$$K_{\vec{U}} = \begin{bmatrix} K_{\vec{X}} & K_{\vec{X}\vec{Y}} \\ K_{\vec{X}\vec{Y}}^T & K_{\vec{Y}} \end{bmatrix} \quad ; \quad K_{\vec{U}}^{-1} = \begin{bmatrix} B & C \\ C^T & D \end{bmatrix} \quad (2.25)$$

Here $K_{\vec{X}} = E[\vec{X}\vec{X}^T]$, $K_{\vec{X}\vec{Y}} = E[\vec{X}\vec{Y}^T]$, and $K_{\vec{Y}} = E[\vec{Y}\vec{Y}^T]$. The blocks $K_{\vec{X}}$, $K_{\vec{Y}}$, B , and D are all non-singular (see Exercise 2.11). We will evaluate the blocks B , C , and D in terms of $K_{\vec{X}}$, $K_{\vec{Y}}$, and $K_{\vec{X}\vec{Y}}$ later, but first we find the conditional density, $p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y})$ in terms of these blocks. What we shall find is that for any given \vec{y} , this is a Gaussian density with a conditional covariance matrix equal to B^{-1} . As in (2.24), where X and Y are one-dimensional, this covariance does not depend on \vec{y} . Also, the conditional mean of \vec{X} , given $\vec{Y} = \vec{y}$, will turn out to be $-B^{-1}C\vec{y}$. Thus, \vec{X} , conditional on $\vec{Y} = \vec{y}$, is $\mathcal{N}(-B^{-1}C\vec{y}, B^{-1})$, i.e.,

$$p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y}) = \frac{\exp \left[-\frac{1}{2}(\vec{x} + B^{-1}C\vec{y})^T B (\vec{x} + B^{-1}C\vec{y}) \right]}{(2\pi)^{n/2} \sqrt{\det(B^{-1})}} \quad (2.26)$$

In order to see this, express $p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y})$ as $p_{\vec{X}\vec{Y}}(\vec{x}, \vec{y}) / p_{\vec{Y}}(\vec{y})$. From (2.17),

$$\begin{aligned} p_{\vec{X},\vec{Y}}(\vec{x}, \vec{y}) &= \frac{\exp \left[-\frac{1}{2}(\vec{x}^T, \vec{y}^T) K_{\vec{U}}^{-1} (\vec{x}^T, \vec{y}^T)^T \right]}{(2\pi)^{(n+m)/2} \sqrt{\det(K_{\vec{U}}^{-1})}} \\ &= \frac{\exp \left[-\frac{1}{2} \left(\vec{x}^T B \vec{x} + \vec{x}^T C \vec{y} + \vec{y}^T C^T \vec{x} + \vec{y}^T D \vec{y} \right) \right]}{(2\pi)^{(n+m)/2} \sqrt{\det(K_{\vec{U}}^{-1})}} \end{aligned}$$

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

$$p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y}) = f(\vec{y}) \exp \left\{ -\frac{\left[\vec{x}^T B \vec{x} + \vec{x}^T C \vec{y} + \vec{y}^T C^T \vec{x} \right]}{2} \right\} \quad (2.27)$$

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

$$p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y}) = f(\vec{y}) \exp \left[\frac{-\left(\vec{x} + B^{-1}C\vec{y} \right)^T B \left(\vec{x} + B^{-1}C\vec{y} \right) + \vec{y}^T C^T B^{-1}C\vec{y}}{2} \right]$$

⁴Exercise 2.12 generalizes this to the case with an arbitrary mean.

Since the last term in the exponent does not depend on \vec{x} , we can absorb it into $f(\vec{y})$. The remaining expression is in the form of (2.26). Since $p_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y})$ must be a probability density for each \vec{y} , the modified coefficient $f(\vec{y})$ must be the reciprocal of the denominator in (2.26), so that $p_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y})$ is given by (2.26).

To interpret (2.26), note that for any sample value \vec{y} for \vec{Y} , the conditional distribution of \vec{X} has a mean given by $-B^{-1}C\vec{y}$ and a Gaussian fluctuation around the mean of variance B^{-1} . This fluctuation has the same distribution for all \vec{y} and thus can be represented as a $\vec{r}\vec{V}$ that is independent of \vec{Y} . Thus we can represent \vec{X} as

$$\vec{X} = G\vec{Y} + \vec{V}; \quad \vec{Y}, \vec{V} \text{ independent} \quad (2.28)$$

where

$$G = -B^{-1}C \quad \text{and} \quad \vec{V} \sim \mathcal{N}(0, B^{-1}) \quad (2.29)$$

\vec{V} is often called an *innovation* because it is the part of \vec{X} that is independent of \vec{Y} . It is also called a *noise term* for the same reason. We will call $K_{\vec{V}} = B^{-1}$ the *conditional covariance* of \vec{X} given any sample value \vec{y} for \vec{Y} . In summary, the unconditional covariance, $K_{\vec{X}}$, of \vec{X} is given by the upper left block of $K_{\vec{U}}$ in (2.25), while the conditional covariance $K_{\vec{V}}$ is the inverse of the upper left block, B , of the inverse of $K_{\vec{U}}$.

From (2.28), we can express G and $K_{\vec{V}}$ in terms of the covariances of \vec{X} and \vec{Y} . To do this, note that

$$K_{\vec{X}\vec{Y}} = E[\vec{X}\vec{Y}^T] = E[G\vec{Y}\vec{Y}^T + \vec{V}\vec{Y}^T] = GK_{\vec{Y}} \quad (2.30)$$

$$K_{\vec{X}} = E[(G\vec{Y} + \vec{V})(G\vec{Y} + \vec{V})^T] = GK_{\vec{Y}}G^T + K_{\vec{V}} \quad (2.31)$$

Solving these equations, we get

$$G = K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1} \quad (2.32)$$

$$K_{\vec{V}} = K_{\vec{X}} - GK_{\vec{Y}}G^T = K_{\vec{X}} - K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}K_{\vec{X}\vec{Y}}^T \quad (2.33)$$

where we have substituted the solution from (2.32) into (2.33). We can summarize these results in the following theorem.

Theorem 2.3 Let \vec{X} and \vec{Y} be zero mean, jointly Gaussian, and jointly non-singular. Then they can be represented in the form $\vec{X} = G\vec{Y} + \vec{V}$ where \vec{Y} and \vec{V} are independent, $\vec{V} \sim \mathcal{N}(0, K_{\vec{V}})$, G is given by (2.32), $K_{\vec{V}}$ is non-singular and given by (2.33), and the conditional density of \vec{X} given \vec{Y} is

$$p_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y}) = \frac{\exp\left[-\frac{1}{2}(\vec{x} - G\vec{y})^T K_{\vec{V}}^{-1}(\vec{x} - G\vec{y})\right]}{(2\pi)^{n/2} \sqrt{\det(K_{\vec{V}})}} \quad (2.34)$$

We can also solve for the matrix $B = K_{\vec{V}}^{-1}$ from (2.33),

$$B = \left[K_{\vec{X}} - K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}K_{\vec{X}\vec{Y}}^T\right]^{-1} \quad (2.35)$$

Similarly, from (2.29) and (2.32),

$$C = -K_{\vec{V}}^{-1}G = -K_{\vec{V}}^{-1}K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1} \quad (2.36)$$

From the symmetry between \vec{X} and \vec{Y} , we can repeat the above arguments, reversing the roles of \vec{X} and \vec{Y} . Thus, we can represent \vec{Y} as

$$\vec{Y} = H\vec{X} + \vec{Z}; \quad \vec{X}, \vec{Z} \text{ independent} \quad (2.37)$$

where

$$H = -D^{-1}C^T \quad \text{and} \quad \vec{Z} \sim \mathcal{N}(0, D^{-1}) \quad (2.38)$$

Using (2.37), we express H and $K_{\vec{Z}}$ in terms of the covariances of \vec{X} and \vec{Y} .

$$K_{\vec{X}\vec{Y}} = E[\vec{X}(H\vec{X} + \vec{Z})^T] = K_{\vec{X}}H^T \quad (2.39)$$

$$K_{\vec{Y}} = E[(H\vec{X} + \vec{Z})(H\vec{X} + \vec{Z})^T] = HK_{\vec{X}}H^T + K_{\vec{Z}} \quad (2.40)$$

Solving, we get

$$H = K_{\vec{X}\vec{Y}}^T K_{\vec{X}}^{-1} \quad (2.41)$$

$$K_{\vec{Z}} = K_{\vec{Y}} - HK_{\vec{X}}H^T = K_{\vec{Y}} - K_{\vec{X}\vec{Y}}^T K_{\vec{X}}^{-1} K_{\vec{X}\vec{Y}} \quad (2.42)$$

The conditional density can now be expressed as

$$p_{\vec{Y}|\vec{X}}(\vec{y}|\vec{x}) = \frac{\exp\left[-\frac{1}{2}(\vec{y} - H\vec{x})^T K_{\vec{Z}}^{-1}(\vec{y} - H\vec{x})\right]}{(2\pi)^{n/2} \sqrt{\det(K_{\vec{Z}})}} \quad (2.43)$$

where H and $K_{\vec{Z}}$ are given in (2.41) and (2.42).

Finally, from (2.42), $D = K_{\vec{Z}}^{-1}$ is

$$D = \left[K_{\vec{Y}} - K_{\vec{X}\vec{Y}}^T K_{\vec{X}}^{-1} K_{\vec{X}\vec{Y}}\right]^{-1} \quad (2.44)$$

Similarly, from (2.38) and (2.41),

$$C^T = -K_{\vec{Z}}^{-1}H = -K_{\vec{Z}}^{-1}K_{\vec{X}\vec{Y}}^T K_{\vec{X}}^{-1} \quad (2.45)$$

The matrices B , C , and D can also be derived directly from setting $K_{\vec{U}}K_{\vec{U}}^{-1} = I$ in (2.26) (see Exercise 2.16).

In many estimation problems, we start with the formulation $\vec{Y} = H\vec{X} + \vec{Z}$ and want to find G and $K_{\vec{V}}$ in the dual formulation $\vec{X} = G\vec{Y} + \vec{V}$. We can use (2.32) and (2.33) directly, but the following expressions are frequently much more useful. From (2.29), we have $G = -K_{\vec{V}}C$, and from (2.38), we have $H = -K_{\vec{Z}}C^T$, or equivalently, $C = -H^T K_{\vec{Z}}^{-1}$. Combining, we have

$$G = K_{\vec{V}}H^T K_{\vec{Z}}^{-1} \quad (2.46)$$

Next, we can multiply $K_{\vec{V}}$, as given by (2.33), by $K_{\vec{V}}^{-1}$ to get

$$I = K_{\vec{X}} K_{\vec{V}}^{-1} - K_{\vec{X}\vec{Y}} K_{\vec{Y}}^{-1} K_{\vec{X}\vec{Y}}^T K_{\vec{V}}^{-1} \quad (2.47)$$

From (2.36) and (2.45), we can express C^T in the following two ways:

$$C^T = -K_{\vec{Y}}^{-1} K_{\vec{X}\vec{Y}}^T K_{\vec{V}}^{-1} = -K_{\vec{Z}}^{-1} H \quad (2.48)$$

The first of these expressions appears at the end of (2.47), and replacing it with the second, (2.47) becomes

$$I = K_{\vec{X}} K_{\vec{V}}^{-1} - K_{\vec{X}\vec{Y}} K_{\vec{Z}}^{-1} H \quad (2.49)$$

Pre-multiplying all terms by $K_{\vec{X}}^{-1}$ and substituting H^T for $K_{\vec{X}}^{-1} K_{\vec{X}\vec{Y}}$, we get the final result,

$$K_{\vec{V}}^{-1} = K_{\vec{X}}^{-1} + H^T K_{\vec{Z}}^{-1} H \quad (2.50)$$

We will interpret these equations in Chapter 4 on estimation.

2.8 Summary

A zero mean n -rv \vec{Z} with a covariance matrix $K_{\vec{Z}}$ is a Gaussian n -rv (equivalently, the n components are jointly Gaussian) if $\vec{b}^T \vec{Z}$ is Gaussian for all real n -vectors \vec{b} . An equivalent condition is that \vec{Z} has the MGF in (2.11). Another equivalent condition is that \vec{Z} is a linear transformation of n IID normalized Gaussian rv's. For the case in which $K_{\vec{Z}}$ is non-singular, a final equivalent condition is that \vec{Z} has the density in (2.17). If \vec{Z} has a singular covariance matrix, it should be viewed as a k -dimensional rv, $k < n$, with a non-singular covariance matrix, plus $n - k$ variables that are linear combinations of the first k . If \vec{Z} has a mean \vec{m} , it is a Gaussian n -rv iff the fluctuation $\vec{Z} - \vec{m}$ is a zero mean Gaussian n -rv.

For a Gaussian n -rv, the regions of equal probability density form concentric ellipsoids around the mean. The principal axes for these ellipsoids are the eigenvectors of the covariance matrix and the length of each axis is proportional to the square root of the corresponding eigenvalue.

If $X_1, X_2, \dots, X_m, Y_1, \dots, Y_n$ are zero mean, jointly Gaussian, and have a non-singular over-all covariance matrix, then the conditional density $p_{\vec{X}|\vec{Y}}(\vec{x} | \vec{y})$ is a Gaussian m -rv for each \vec{y} , and has a fixed covariance $K_{\vec{X}} - K_{\vec{X}\vec{Y}} K_{\vec{Y}}^{-1} K_{\vec{X}\vec{Y}}^T$ for each \vec{y} and has a mean $K_{\vec{X}\vec{Y}} K_{\vec{Y}}^{-1} \vec{y}$, which depends linearly on \vec{y} . This situation can be equivalently formulated as $\vec{X} = G\vec{Y} + \vec{V}$, where \vec{V} is a zero mean Gaussian m -rv independent of \vec{Y} . Equivalently it can be formulated as $\vec{Y} = H\vec{X} + \vec{Z}$ where \vec{Z} is a zero mean Gaussian n -rv independent of \vec{X} .

2.9 Exercises

Exercise 2.1 a) Let X, Y be IID rv's, each with density $p_X(x) = \alpha \exp(-x^2/2)$. In part (b), we show that α must be $1/\sqrt{2\pi}$ in order for $p_X(x)$ 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 $E[X] = 0$ and that $E[X^2] = 1$.

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

Exercise 2.2 a) By expanding in a power series in $(1/2)s^2\sigma^2$, show that

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

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

$$g_Z(s) = E[e^{sZ}] = 1 + sE[Z] + \frac{s^2E[Z^2]}{2} + \cdots + \frac{s^iE[Z^i]}{(i)!} + \cdots$$

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

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

Exercise 2.3 Let X and Z be IID normalized Gaussian random variables. Let $Y = |Z| \text{Sgn}(X)$, where $\text{Sgn}(X)$ 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 2.4 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.

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

Exercise 2.5 a) Let \vec{U} be a rv' with mean \vec{m} and covariance K whose MGF is given by (2.12). Let $X = \vec{s}^T \vec{U}$ for an arbitrary real vector \vec{s} . Show that the MGF of X is given by $g_X(r) = \exp[rE[X] + r^2\sigma_X^2/2]$ and relate $E[X]$ and σ_X^2 to \vec{m} and K .

b) Show that \vec{U} is a Gaussian rv'.

Exercise 2.6 a) Let $\vec{Z} \sim \mathcal{N}(0, K)$ be n -dimensional. By expanding in a power series in $(1/2)\vec{s}^T K \vec{s}$, show that

$$g_{\vec{Z}}(\vec{s}) = \exp \left[\frac{\vec{s}^T K \vec{s}}{2} \right] = 1 + \frac{\sum_{i,j} s_i s_j [K]_{i,j}}{2} + \cdots + \frac{\left(\sum_{i,j} s_i s_j [K]_{i,j} \right)^m}{2^m m!} + \cdots$$

b) By expanding $e^{s_i Z_i}$ in a power series in $s_i Z_i$ for each i , show that

$$g_{\vec{Z}}(s) = E \left[\exp \left(\sum_i s_i Z_i \right) \right] = \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} \frac{s_1^{j_1}}{(j_1)!} \cdots \frac{s_n^{j_n}}{(j_n)!} E[Z_1^{j_1} \cdots Z_n^{j_n}]$$

c) Let $D = \{i_1, i_2, \dots, i_{2m}\}$ be a set of $2m$ distinct integers each between 1 and n . Consider the term $s_{i_1} s_{i_2} \cdots s_{i_{2m}} E[Z_{i_1} Z_{i_2} \cdots Z_{i_{2m}}]$ in part (b). By comparing with the set of terms in part (a) containing the same product $s_{i_1} s_{i_2} \cdots s_{i_{2m}}$, show that

$$E[Z_{i_1} Z_{i_2} \cdots Z_{i_{2m}}] = \sum_{j_1 j_2 \cdots j_{2m}} \frac{[K]_{j_1 j_2} [K]_{j_3 j_4} \cdots [K]_{j_{2m-1} j_{2m}}}{2^m m!}$$

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

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

$$E[Z_{i_1} Z_{i_2} \cdots Z_{i_{2m}}] = \sum_{j_1, j_2, \dots, j_{2m}} [K]_{j_1 j_2} [K]_{j_3 j_4} \cdots [K]_{j_{2m-1} j_{2m}} \quad (2.51)$$

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 $j_{2i-1} < j_{2i}$ for $1 \leq i \leq m$ and $j_{2i-1} < j_{2i+1}$ for $1 \leq i \leq m-1$.

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

Exercise 2.7 Let Q be an orthonormal matrix. Show that the squared distance between any two vectors \vec{z} and \vec{y} is equal to the squared distance between $Q\vec{z}$ and $Q\vec{y}$.

Exercise 2.8 a) Let $K = \begin{bmatrix} .75 & .25 \\ .25 & .75 \end{bmatrix}$. Show that 1 and $1/2$ are eigenvalues of K and find the normalized eigenvectors. Express K as $Q\Lambda Q^{-1}$ where Λ 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) Consider the m^{th} power of K , K^m for $m > 0$. Find the eigenvalues and eigenvectors of K^m .

Exercise 2.9 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 $p_{X|Y}(x|y)$.

Exercise 2.10 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 $p_{X|V}(x|v)$. Hint: This requires no computation.

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

Exercise 2.11 a) Let $\vec{U}^T = (\vec{X}^T, \vec{Y}^T)$ have a non-singular covariance matrix $K_{\vec{U}}$. Show that $K_{\vec{X}}$ and $K_{\vec{Y}}$ are positive definite, and thus non-singular.

b) Show that the matrices B and D in (2.25) are also positive definite and thus non-singular.

Exercise 2.12 Let \vec{X} and \vec{Y} be jointly Gaussian r.v.'s with means $\vec{m}_{\vec{X}}$ and $\vec{m}_{\vec{Y}}$, covariance matrices $K_{\vec{X}}$ and $K_{\vec{Y}}$, and cross covariance matrix $K_{\vec{X}\vec{Y}}$. Find the conditional probability density $p_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y})$. Assume that the covariance of (\vec{X}^T, \vec{Y}^T) is non-singular. Hint: think of the fluctuations of \vec{X} and \vec{Y} .

Exercise 2.13 a) Let \vec{W} be a normalized IID Gaussian n -r.v. and let \vec{Y} be a Gaussian m -r.v. Suppose we would like the cross covariance $E[\vec{W}\vec{Y}^T]$ to be some arbitrary real valued n by m matrix K . Find the matrix A such that $\vec{Y} = A\vec{W}$ achieves the desired cross covariance. Note: this shows that any real valued n by m matrix is the cross covariance matrix for some choice of random vectors.

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

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

Exercise 2.14 a) Let $\vec{W} = (W_1, W_2, \dots, W_{2n})^T$ be a $2n$ dimensional IID normalized Gaussian r.v. Let $S_{2n} = W_1^2 + W_2^2 + \dots + W_{2n}^2$. Show that S_{2n} is an n th order Erlang rv with parameter $1/2$, i.e., that $p_{S_{2n}}(s) = 2^{-n}s^{n-1}e^{-s/2}/(n-1)!$. Hint: look at S_2 from Exercise 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

$$p_{R_{2n}}(r) = b_{2n}(r) p_{\vec{W}}(\vec{w}) \big|_{\vec{W} \cdot \vec{W}^T \vec{W} = r^2}$$

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 2.15 a) Let $\vec{U} = (X, Y)^T$. Solve directly for B , C , and D in (2.25) for this case, and show that (2.26) agrees with (2.24)

b) Show that your solution for B , C , and D agrees with (2.35), (2.36), and (2.44).

Exercise 2.16 Express B , C , and D in terms of $K_{\vec{X}}$, $K_{\vec{Y}}$ and $K_{\vec{X}\vec{Y}}$ by multiplying the block expression for $K_{\vec{U}}$ by that for $K_{\vec{U}}^{-1}$. Show that your answer agrees with that in (2.35), (2.36), and (2.44).

