CHAPTER 2

Random Vectors

The basic object of study in this book is the random vector and its induced distribution in an inner product space. Here, utilizing the results outlined in Chapter 1, we introduce random vectors, mean vectors, and covariances. Characteristic functions are discussed and used to give the well known factorization criterion for the independence of random vectors. Two special classes of distributions, the orthogonally invariant distributions and the weakly spherical distributions, are used for illustrative purposes. The vector spaces that occur in this chapter are all finite dimensional.

2.1. RANDOM VECTORS

Before a random vector can be defined, it is necessary to first introduce the Borel sets of a finite dimensional inner product space $(V, (\cdot, \cdot))$. Setting $\|x\| = (x, x)^{1/2}$, the open ball of radius r about x_0 is the set defined by $S_r(x_0) \equiv \{x \mid \|x - x_0\| < r\}$.

Definition 2.1. The Borel σ-algebra of $(V, (\cdot, \cdot))$, denoted by $\mathcal{B}(V)$, is the smallest σ-algebra that contains all of the open balls.

Since any two inner products on V are related by a positive definite linear transformation, it follows that $\mathcal{B}(V)$ does not depend on the inner product on V—that is, if we start with two inner products on V and use these inner products to generate a Borel σ-algebra, the two σ-algebras are the same. Thus we simply call $\mathcal{B}(V)$ the Borel σ-algebra of V without mentioning the inner product.

A probability space is a triple $(\Omega, \mathcal{F}, P_0)$ where Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P_0 is a probability measure defined on $\mathcal{F}$.

Definition 2.2. A random vector $X \in V$ is a function mapping Ω into V such that $X^{-1}(B) \in \mathcal{F}$ for each Borel set $B \in \mathcal{B}(V)$. Here, $X^{-1}(B)$ is the inverse image of the set B.

Since the space on which a random vector is defined is usually not of interest here, the argument of a random vector X is ordinarily suppressed. Further, it is the induced distribution of X on V that most interests us. To define this, consider a random vector X defined on Ω to V where $(\Omega, \mathcal{F}, P_0)$ is a probability space. For each Borel set $B \in \mathcal{B}(V)$, let $Q(B) = P_0(X^{-1}(B))$. Clearly, Q is a probability measure on $\mathcal{B}(V)$ and Q is called the *induced distribution* of X—that is, Q is induced by X and P_0. The following result shows that any probability measure Q on $\mathcal{B}(V)$ is the induced distribution of some random vector.

Proposition 2.1. Let Q be a probability measure on $\mathcal{B}(V)$ where V is a finite dimensional inner product space. Then there exists a probability space $(\Omega, \mathcal{F}, P_0)$ and a random vector X on Ω to V such that Q is the induced distribution of X.

Proof. Take $\Omega = V$, $\mathcal{F} = \mathcal{B}(V)$, $P_0 = Q$, and let $X(\omega) = \omega$ for $\omega \in V$. Clearly, the induced distribution of X is Q. □

Henceforth, we write things like: "Let X be a random vector in V with distribution Q," to mean that X is a random vector and its induced distribution is Q. Alternatively, the notation $\mathcal{L}(X) = Q$ is also used—this is read: "The distributional law of X is Q."

A function f defined on V to W is called *Borel measurable* if the inverse image of each set $B \in \mathcal{B}(W)$ is in $\mathcal{B}(V)$. Of course, if X is a random vector in V, then $f(X)$ is a random vector in W when f is Borel measurable. In particular, when f is continuous, f is Borel measurable. If $W = R$ and f is Borel measurable on V to R, then $f(X)$ is a real-valued random variable.

Definition 2.3. Suppose X is a random vector in V with distribution Q and f is a real-valued Borel measurable function defined on V. If $\int_V |f(x)| Q(dx) < +\infty$, then we say that $f(X)$ has *finite expectation* and we write $\mathcal{E}f(X)$ for $\int_V f(x) Q(dx)$.

In the above definition and throughout this book, all integrals are Lebesgue integrals, and all functions are assumed Borel measurable.

◆ **Example 2.1.** Take V to be the coordinate space R^n with the usual inner product $(\cdot, \cdot)$ and let dx denote standard Lebesgue measure on R^n. If q is a non-negative function on R^n such that $\int q(x)\,dx = 1$, then q is called a *density function*. It is clear that the measure Q given by $Q(B) = \int_B q(x)\,dx$ is a probability measure on R^n so Q is the distribution of some random vector X on R^n. If $\varepsilon_1, \ldots, \varepsilon_n$ is the standard basis for R^n, then $(\varepsilon_i, X) \equiv X_i$ is the *ith coordinate* of X. Assume that X_i has a finite expectation for $i = 1, \ldots, n$. Then $\mathcal{E}X_i = \int_{R^n}(\varepsilon_i, x)q(x)\,dx \equiv \mu_i$ is called the *mean value* of X_i and the vector $\mu \in R^n$ with coordinates $\mu_1, \ldots, \mu_n$ is the *mean vector* of X. Notice that for any vector $x \in R^n$, $\mathcal{E}(x, X) = \mathcal{E}(\Sigma x_i \varepsilon_i, X) = \Sigma x_i \mathcal{E}(\varepsilon_i, X) = \Sigma x_i \mu_i = (x, \mu)$. Thus the mean vector μ satisfies the equation $\mathcal{E}(x, X) = (x, \mu)$ for all $x \in R^n$ and μ is clearly unique. It is exactly this property of μ that we use to define the mean vector of a random vector in an arbitrary inner product space V. ◆

Suppose X is a random vector in an inner product space $(V, (\cdot, \cdot))$ and assume that for each $x \in V$, the random variable (x, X) has a finite expectation. Let $f(x) = \mathcal{E}(x, X)$, so f is a real-valued function defined on V. Also, $f(\alpha_1 x_1 + \alpha_2 x_2) = \mathcal{E}(\alpha_1 x_1 + \alpha_2 x_2, X) = \mathcal{E}[\alpha_1(x_1, X) + \alpha_2(x_2, X)] = \alpha_1 \mathcal{E}(x_1, X) + \alpha_2 \mathcal{E}(x_2, X) = \alpha_1 f(x_1) + \alpha_2 f(x_2)$. Thus f is a linear function on V. Therefore, there exists a unique vector $\mu \in V$ such that $f(x) = (x, \mu)$ for all $x \in V$. Summarizing, there exists a unique vector $\mu \in V$ that satisfies $\mathcal{E}(x, X) = (x, \mu)$ for all $x \in V$. The vector μ is called the *mean vector* of X and is denoted by $\mathcal{E}X$. This notation leads to the suggestive equation $\mathcal{E}(x, X) = (x, \mathcal{E}X)$, which we know is valid in the coordinate case.

Proposition 2.2. Suppose $X \in (V, (\cdot, \cdot))$ and assume X has a mean vector μ. Let $(W, [\cdot, \cdot])$ be an inner product space and consider $A \in \mathcal{L}(V, W)$ and $w_0 \in W$. Then the random vector $Y = AX + w_0$ has the mean vector $A\mu + w_0$—that is, $\mathcal{E}Y = A\mathcal{E}X + w_0$.

Proof. The proof is a computation. For $w \in W$,

$$\begin{aligned}\mathcal{E}[w, Y] &= \mathcal{E}[w, AX + w_0] = \mathcal{E}[w, AX] + [w, w_0] \\ &= \mathcal{E}(A'w, X) + [w, w_0] = (A'w, \mu) + [w, w_0] \\ &= [w, A\mu] + [w, w_0] = [w, A\mu + w_0].\end{aligned}$$

Thus $A\mu + w_0$ satisfies the defining equation for the mean vector of Y and by the uniqueness of mean vectors, $EY = A\mu + w_0$. □

If X_1 and X_2 are both random vectors in $(V, (\cdot, \cdot))$, which have mean vectors, then it is easy to show that $\mathcal{E}(X_1 + X_2) = \mathcal{E}X_1 + \mathcal{E}X_2$. The following proposition shows that the mean vector μ of a random vector does not depend on the inner product on V.

Proposition 2.3. If X is a random vector in $(V, (\cdot, \cdot))$ with mean vector μ satisfying $\mathcal{E}(x, X) = (x, \mu)$ for all $x \in V$, then μ satisfies $\mathcal{E}f(x, X) = f(x, \mu)$ for every bilinear function f on $V \times V$.

Proof. Every bilinear function f is given by $f(x_1, x_2) = (x_1, Ax_2)$ for some $A \in \mathcal{L}(V, V)$. Thus $\mathcal{E}f(x, X) = \mathcal{E}(x, AX) = (x, A\mu) = f(x, \mu)$ where the second equality follows from Proposition 2.2. □

When the bilinear function f is an inner product on V, the above result establishes that the mean vector is inner product free. At times, a convenient choice of an inner product can simplify the calculation of a mean vector.

The definition and basic properties of the covariance between two real-valued random variables were covered in Example 1.9. Before defining the covariance of a random vector, a review of covariance matrices for coordinate random vectors in R^n is in order.

◆ **Example 2.2.** In the notation of Example 2.1, consider a random vector X in R^n with coordinates $X_i = (\varepsilon_i, X)$ where $\varepsilon_1, \ldots, \varepsilon_n$ is the standard basis for R^n and $(\cdot, \cdot)$ is the standard inner product. Assume that $\mathcal{E}X_i^2 < +\infty$, $i = 1, \ldots, n$. Then $\text{cov}(X_i, X_j) \equiv \sigma_{ij}$ exists for all $i, j = 1, \ldots, n$. Let Σ be the $n \times n$ matrix with elements σ_{ij}. Of course, σ_{ii} is the variance of X_i and σ_{ij} is the covariance between X_i and X_j. The symmetric matrix Σ is called the *covariance matrix* of X. Consider vectors $x, y \in R^n$ with coordinates x_i and y_i, $i = 1, \ldots, n$. Then

$$\begin{aligned}
\text{cov}\{(x, X), (y, X)\} &= \text{cov}\left\{\sum_i x_i X_i, \sum_j y_j X_j\right\} \\
&= \sum_i \sum_j x_i y_j \,\text{cov}(X_i, X_j) = \sum_i \sum_j x_i y_j \sigma_{ij} \\
&= (x, \Sigma y).
\end{aligned}$$

Hence $\text{cov}\{(x, X), (y, X)\} = (x, \Sigma y)$. It is this property of Σ that is used to define the covariance of a random vector. ◆

With the above example in mind, consider a random vector X in an inner product space $(V, (\cdot, \cdot))$ and assume that $\mathcal{E}(x, X)^2 < \infty$ for all $x \in V$. Thus

(x, X) has a finite variance and the covariance between (x, X) and (y, X) is well defined for each $x, y \in V$.

Proposition 2.4. For $x, y \in V$, define $f(x, y)$ by

$$f(x, y) = \text{cov}\{(x, X), (y, X)\}.$$

Then f is a non-negative definite bilinear function on $V \times V$.

Proof. Clearly, $f(x, y) = f(y, x)$ and $f(x, x) = \text{var}\{(x, X)\} \geqslant 0$, so it remains to show that f is bilinear. Since f is symmetric, it suffices to verify that $f(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 f(x_1, y) + \alpha_2 f(x_2, y)$. This verification goes as follows:

$$\begin{aligned} f(\alpha_1 x_1 + \alpha_2 x_2, y) &= \text{cov}\{(\alpha_1 x_1 + \alpha_2 x_2, X), (y, X)\} \\ &= \text{cov}\{\alpha_1(x_1, X) + \alpha_2(x_2, X), (y, X)\} \\ &= \alpha_1 \text{cov}\{(x_1, X), (y, X)\} + \alpha_2 \text{cov}\{(x_2, X), (y, X)\} \\ &= \alpha_1 f(x_1, y) + \alpha_2 f(x_2, y). \end{aligned}$$

□

By Proposition 1.26, there exists a unique non-negative definite linear transformation Σ such that $f(x, y) = (x, \Sigma y)$.

Definition 2.4. The unique non-negative definite linear transformation Σ on V to V that satisfies

$$\text{cov}\{(x, X), (y, X)\} = (x, \Sigma y)$$

is called the *covariance* of X and is denoted by $\text{Cov}(X)$.

Implicit in the above definition is the assumption that $\mathcal{E}(x, X)^2 < +\infty$ for all $x \in V$. Whenever we discuss covariances of random vectors, $\mathcal{E}(x, X)^2$ is always assumed finite.

It should be emphasized that the covariance of a random vector in $(V, (\cdot, \cdot))$ depends on the given inner product. The next result shows how the covariance changes as a function of the inner product.

Proposition 2.5. Consider a random vector X in $(V, (\cdot, \cdot))$ and suppose $\text{Cov}(X) = \Sigma$. Let $[\cdot, \cdot]$ be another inner product on V given by $[x, y] = (x, Ay)$ where A is positive definite on $(V, (\cdot, \cdot))$. Then the covariance of X in the inner product space $(V, [\cdot, \cdot])$ is ΣA.

Proof. To verify that ΣA is the covariance for X in $(V, [\cdot, \cdot])$, we must show that $\text{cov}\{[x, X], [y, X]\} = [x, \Sigma Ay]$ for all $x, y \in V$. To do this, use the definition of $[\cdot, \cdot]$ and compute:

$$\begin{aligned}\text{cov}\{[x, X], [y, X]\} &= \text{cov}\{(x, AX), (y, AX)\} = \text{cov}\{(Ax, X), (Ay, X)\} \\ &= (Ax, \Sigma Ay) = (x, A\Sigma Ay) = [x, \Sigma Ay].\end{aligned}$$ □

Two immediate consequences of Proposition 2.5 are: (i) if $\text{Cov}(X)$ exists in one inner product, then it exists in all inner products, and (ii) if $\text{Cov}(X) = \Sigma$ in $(V, (\cdot, \cdot))$ and if Σ is positive definite, then the covariance of X in the inner product $[x, y] \equiv (x, \Sigma^{-1}y)$ is the identity linear transformation. The result below often simplifies a computation involving the derivation of a covariance.

Proposition 2.6. Suppose $\text{Cov}(X) = \Sigma$ in $(V, (\cdot, \cdot))$. If Σ_1 is a self-adjoint linear transformation on $(V, (\cdot, \cdot))$ to $(V, (\cdot, \cdot))$ that satisfies

(2.1) $$\text{var}\{(x, X)\} = (x, \Sigma_1 x) \qquad \text{for } x \in V,$$

then $\Sigma_1 = \Sigma$.

Proof. Equation (2.1) implies that $(x, \Sigma_1 x) = (x, \Sigma x)$, $x \in V$. Since Σ_1 and Σ are self-adjoint, Proposition 1.16 yields the conclusion $\Sigma_1 = \Sigma$. □

When $\text{Cov}(X) = \Sigma$ is singular, then the random vector X takes values in the translate of a subspace of $(V, (\cdot, \cdot))$. To make this precise, let us consider the following.

Proposition 2.7. Let X be a random vector in $(V, (\cdot, \cdot))$ and suppose $\text{Cov}(X) = \Sigma$ exists. With $\mu = \mathcal{E}X$ and $\mathcal{R}(\Sigma)$ denoting the range of Σ, $P\{X \in \mathcal{R}(\Sigma) + \mu\} = 1$.

Proof. The set $\mathcal{R}(\Sigma) + \mu$ is the set of vectors of the form $x + \mu$ for $x \in \mathcal{R}(\Sigma)$; that is $\mathcal{R}(\Sigma) + \mu$ is the translate, by μ, of the subspace $\mathcal{R}(\Sigma)$. The statement $P\{X \in \mathcal{R}(\Sigma) + \mu\} = 1$ is equivalent to the statement $P\{X - \mu \in \mathcal{R}(\Sigma)\} = 1$. The random vector $Y = X - \mu$ has mean zero and, by Proposition 2.6, $\text{Cov}(Y) = \text{Cov}(X) = \Sigma$ since $\text{var}\{(x, X - \mu)\} = \text{var}\{(x, X)\}$ for $x \in V$. Thus it must be shown that $P\{Y \in \mathcal{R}(\Sigma)\} = 1$. If Σ is nonsingular, then $\mathcal{R}(\Sigma) = V$ and there is nothing to show. Thus assume that the null space of Σ, $\mathcal{N}(\Sigma)$, has dimension $k > 0$ and let $\{x_1, \ldots, x_k\}$ be an orthonormal basis for $\mathcal{N}(\Sigma)$. Since $\mathcal{R}(\Sigma)$ and $\mathcal{N}(\Sigma)$ are perpendicular and $\mathcal{R}(\Sigma) \oplus$

$\mathfrak{N}(\Sigma) = V$, a vector x is not in $\mathfrak{R}(\Sigma)$ iff for some index $i = 1, \ldots, k$, $(x_i, x) \neq 0$. Thus

$$P\{Y \notin \mathfrak{R}(\Sigma)\} = P\{(x_i, Y) \neq 0 \text{ for some } i = 1, \ldots, k\}$$
$$\leqslant \sum_1^k P\{(x_i, Y) \neq 0\}.$$

But (x_i, Y) has mean zero and $\text{var}\{(x_i, Y)\} = (x_i, \Sigma x_i) = 0$ since $x_i \in \mathfrak{N}(\Sigma)$. Thus (x_i, Y) is zero with probability one, so $P\{(x_i, Y) \neq 0\} = 0$. Therefore $P\{Y \notin \mathfrak{R}(\Sigma)\} = 0$. □

Proposition 2.2 describes how the mean vector changes under linear transformations. The next result shows what happens to the covariance under linear transformations.

Proposition 2.8. Suppose X is a random vector in $(V, (\cdot, \cdot))$ with $\text{Cov}(X) = \Sigma$. If $A \in \mathcal{L}(V, W)$ where $(W, [\cdot, \cdot])$ is an inner product space, then

$$\text{Cov}(AX + w_0) = A\Sigma A'$$

for all $w_0 \in W$.

Proof. By Proposition 2.6, it suffices to show that for each $w \in W$,

$$\text{var}[w, AX + w_0] = [w, A\Sigma A'w].$$

However,

$$\text{var}[w, AX + w_0] = \text{var}([w, AX] + [w, w_0]) = \text{var}[w, AX]$$
$$= \text{var}(A'w, X) = (A'w, \Sigma A'w) = [w, A\Sigma A'w].$$

Thus $\text{Cov}(AX + w_0) = A\Sigma A'$. □

2.2. INDEPENDENCE OF RANDOM VECTORS

With the basic properties of mean vectors and covariances established, the next topic of discussion is characteristic functions and independence of random vectors. Let X be a random vector in $(V, (\cdot, \cdot))$ with distribution Q.

Definition 2.5. The complex valued function on V defined by

$$\phi(v) \equiv \mathcal{E} e^{i(v, X)} = \int_V e^{i(v, x)} Q(dx)$$

is the *characteristic function* of X.

In the above definition, $e^{it} = \cos t + i \sin t$ where $i = \sqrt{-1}$ and $t \in R$. Since e^{it} is a bounded continuous function of t, characteristic functions are well defined for all distributions Q on $(V, (\cdot, \cdot))$. Forthcoming applications of characteristic functions include the derivation of distributions of certain functions of random vectors and a characterization of the independence of two or more random vectors.

One basic property of characteristic functions is their uniqueness, that is, if Q_1 and Q_2 are probability distributions on $(V, (\cdot, \cdot))$ with characteristic functions ϕ_1 and ϕ_2, and if $\phi_1(x) = \phi_2(x)$ for all $x \in V$, then $Q_1 = Q_2$. A proof of this is based on the multidimensional Fourier inversion formula, which can be found in Cramér (1946). A consequence of this uniqueness is that, if X_1 and X_2 are random vectors in $(V, (\cdot, \cdot))$ such that $\mathcal{L}((x, X_1)) = \mathcal{L}((x, X_2))$ for all $x \in V$, then $\mathcal{L}(X_1) = \mathcal{L}(X_2)$. This follows by observing that $\mathcal{L}((x, X_1)) = \mathcal{L}((x, X_2))$ for all x implies the characteristic functions of X_1 and X_2 are the same and hence their distributions are the same.

To define independence, consider a probability space $(\Omega, \mathcal{F}, P_0)$ and let $X \in (V, (\cdot, \cdot))$ and $Y \in (W, [\cdot, \cdot])$ be two random vectors defined on Ω.

Definition 2.6. The random vectors X and Y are *independent* if for any Borel sets $B_1 \in \mathcal{B}(V)$ and $B_2 \in \mathcal{B}(W)$,

$$P_0\{X^{-1}(B_1) \cap Y^{-1}(B_2)\} = P_0\{X^{-1}(B_1)\} P_0\{Y^{-1}(B_2)\}.$$

In order to describe what independence means in terms of the induced distributions of $X \in (V, (\cdot, \cdot))$ and $Y \in (W, [\cdot, \cdot])$, it is necessary to define what is meant by the joint induced distribution of X and Y. The natural vector space in which to have X and Y take values is the direct sum $V \oplus W$ defined in Chapter 1. For $\{v_i, w_i\} \in V \oplus W$, $i = 1, 2$, define the inner product $(\cdot, \cdot)_1$ by

$$(\{v_1, w_1\}, \{v_2, w_2\})_1 = (v_1, v_2) + [w_1, w_2].$$

That $(\cdot, \cdot)_1$ is an inner product on $V \oplus W$ is routine to check. Thus $\{X, Y\}$ takes values in the inner product space $V \oplus W$. However, it must be shown that $\{X, Y\}$ is a Borel measurable function. Briefly, this argument goes as follows. The space $V \oplus W$ is a Cartesian product space—that is, $V \oplus W$ consists of all pairs $\{v, w\}$ with $v \in V$ and $w \in W$. Thus one way to get a σ-algebra on $V \oplus W$ is to form the product σ-algebra $\mathcal{B}(V) \times \mathcal{B}(W)$, which is the smallest σ-algebra containing all the product Borel sets $B_1 \times B_2 \subseteq V \oplus W$ where $B_1 \in \mathcal{B}(V)$ and $B_2 \in \mathcal{B}(W)$. It is not hard to verify that inverse images, under $\{X, Y\}$, of sets in $\mathcal{B}(V) \times \mathcal{B}(W)$ are in the σ-algebra $\mathcal{F}$. But the product σ-algebra $\mathcal{B}(V) \times \mathcal{B}(W)$ is just the σ-algebra $\mathcal{B}(V \oplus W)$ defined earlier. Thus $\{X, Y\} \in V \oplus W$ is a random vector and hence has an induced distribution Q defined on $\mathcal{B}(V \oplus W)$. In addition, let Q_1 be the induced distribution of X on $\mathcal{B}(V)$ and let Q_2 be the induced distribution of Y on $\mathcal{B}(W)$. It is clear that $Q_1(B_1) = Q(B_1 \times W)$ for $B_1 \in \mathcal{B}(V)$ and $Q_2(B_2) = Q(V \times B_2)$ for $B_2 \in \mathcal{B}(W)$. Also, the characteristic function of $\{X, Y\} \in V \oplus W$ is

$$\phi(\{v, w\}) = \mathcal{E}\exp\left[i(\{v, w\}, \{X, Y\})_1\right] = \mathcal{E}\exp(i(v, X) + i[w, Y])$$

and the marginal characteristic functions of X and Y are

$$\phi_1(v) = \mathcal{E}e^{i(v, X)}$$

and

$$\phi_2(w) = \mathcal{E}e^{i[w, Y]}.$$

Proposition 2.9. Given random vectors $X \in (V, (\cdot, \cdot))$ and $Y \in (W, [\cdot, \cdot])$, the following are equivalent:

(i) X and Y are independent.
(ii) $Q(B_1 \times B_2) = Q_1(B_1)Q_2(B_2)$ for all $B_1 \in \mathcal{B}(V)$ and $B_2 \in \mathcal{B}(W)$.
(iii) $\phi(\{v, w\}) = \phi_1(v)\phi_2(w)$ for all $v \in V$ and $w \in W$.

Proof. By definition,

$$Q(B_1 \times B_2) = P_0\{\{X, Y\} \in B_1 \times B_2\} = P_0\{X \in B_1, Y \in B_2\}.$$

The equivalence of (i) and (ii) follows immediately from the above equation. To show (ii) implies (iii), first note that, if f_1 and f_2 are integrable complex valued functions on V and W, then when (ii) holds,

$$\int_{V \oplus W} f_1(v) f_2(w) Q(dv, dw) = \int_V \int_W f_1(v) f_2(w) Q_1(dv) Q_2(dw)$$

$$= \int_V f_1(v) Q_1(dv) \int_W f_2(w) Q_2(dw)$$

by Fubini's Theorem (see Chung, 1968). Taking $f_1(v) = e^{i(v_1, v)}$ for v_1, $v \in V$, and $f_2(w) = e^{i[w_1, w]}$ for w_1, $w \in W$, we have

$$\begin{aligned}\phi(\{v_1, w_1\}) &= \int \exp(i(v_1, v) + i[w_1, w]) Q(dv, dw) \\ &= \int_V \exp[i(v_1, v)] Q_1(dv) \int_W \exp(i[w_1, w]) Q_2(dw) \\ &= \phi_1(v_1)\phi_2(w_1).\end{aligned}$$

Thus (ii) implies (iii). For (iii) implies (ii), note that the product measure $Q_1 \times Q_2$ has characteristic function $\phi_1\phi_2$. The uniqueness of characteristic functions then implies that $Q = Q_1 \times Q_2$. □

Of course, all of the discussion above extends to the case of more than two random vectors. For completeness, we briefly describe the situation. Given a probability space $(\Omega, \mathcal{F}, P_0)$ and random vectors $X_j \in (V_j, (\cdot,\cdot)_j)$, $j = 1,\dots, k$, let Q_j be the induced distribution of X_j and let ϕ_j be the characteristic function of X_j. The random vectors $X_1,\dots, X_k$ are *independent* if for all $B_j \in \mathfrak{B}(V_j)$,

$$P_0\{X_j \in B_j, j = 1,\dots, k\} = \prod_{j=1}^{k} P_0\{X_j \in B_j\}.$$

To construct one random vector from $X_1,\dots, X_k$, consider the direct sum $V_1 \oplus \cdots \oplus V_k$ with the inner product $(\cdot,\cdot) = \Sigma_1^k(\cdot,\cdot)_j$. In other words, if $\{v_1,\dots, v_k\}$ and $\{w_1,\dots, w_k\}$ are elements of $V_1 \oplus \cdots \oplus V_k$, then the inner product between these vectors is $\Sigma_1^k(v_j, w_j)_j$. An argument analogous to that given earlier shows that $\{X_1,\dots, X_k\}$ is a random vector in $V_1 \oplus \cdots \oplus V_k$ and the Borel σ-algebra of $V_1 \oplus \cdots \oplus V_k$ is just the product σ-algebra $\mathfrak{B}(V_1) \times \cdots \times \mathfrak{B}(V_k)$. If Q denotes the induced distribution of $\{X_1,\dots, X_k\}$, then the independence of $X_1,\dots, X_k$ is equivalent to the assertion that

$$Q(B_1 \times \cdots \times B_k) = \prod_{j=1}^{k} Q_j(B_j)$$

for all $B_j \in \mathfrak{B}(V_j)$, $j = 1,\dots, k$, and this is equivalent to

$$\mathcal{E}\exp\left[i\sum_1^k (v_j, X_j)_j\right] = \prod_{j=1}^{k} \phi_j(v_j).$$

Of course, when $X_1,\dots, X_k$ are independent and f_j is an integrable real valued function on V_j, $j = 1,\dots, k$, then

$$\mathcal{E}\prod_{j=1}^{k} f_j(X_j) = \prod_{j=1}^{k} \mathcal{E}f_j(X_j).$$

This equality follows from the fact that

$$Q(B_1 \times \cdots \times B_k) = \prod_{j=1}^{k} Q_j(B_j)$$

and Fubini's Theorem.

◆ **Example 2.3.** Consider the coordinate space R^p with the usual inner product and let Q_0 be a fixed distribution on R^p. Suppose $X_1, \ldots, X_n$ are independent with each $X_i \in R^p$, $i = 1, \ldots, n$, and $\mathcal{L}(X_i) = Q_0$. That is, there is a probability space $(\Omega, \mathcal{F}, P_0)$, each X_i is a random vector on Ω with values in R^p, and for Borel sets,

$$P_0\{X_i \in B_i, i = 1, \ldots, n\} = \prod_1^n Q_0(B_i).$$

Thus $\{X_1, \ldots, X_n\}$ is a random vector in the direct sum $R^p \oplus \cdots \oplus R^p$ with n terms in the sum. However, there are a variety of ways to think about the above direct sum. One possibility is to form the coordinate random vector

$$Y = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \in R^{np}$$

and simply consider Y as a random vector in R^{np} with the usual inner product. A disadvantage of this representation is that the independence of $X_1, \ldots, X_n$ becomes slightly camouflaged by the notation. An alternative is to form the random matrix

$$X = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix} \in \mathcal{L}_{p,n}.$$

Thus X has rows X_i', $i = 1, \ldots, n$, which are independent and each has distribution Q_0. The inner product on $\mathcal{L}_{p,n}$ is just that inherited from the standard inner products on R^n and R^p. Therefore X is a random vector in the inner product space $(\mathcal{L}_{p,n}, \langle \cdot , \cdot \rangle)$. In the sequel, we ordinarily represent $X_1, \ldots, X_n$ by the random vector $X \in \mathcal{L}_{p,n}$. The advantages of this representation are far from clear at this point, but the reader should be convinced by the end of this book that such a choice is not unreasonable. The derivation of the mean and covariance of $X \in \mathcal{L}_{p,n}$ given in the next section should provide some evidence that the above representation is useful. ◆

2.3. SPECIAL COVARIANCE STRUCTURES

In this section, we derive the covariances of some special random vectors. The orthogonally invariant probability distributions on a vector space are shown to have covariances that are a constant times the identity transformation. In addition, the covariance of the random vector given in Example 2.3 is shown to be a Kronecker product. The final example provides an expression for the covariance of an outer product of a random vector with itself.

Suppose $(V, (\cdot, \cdot))$ is an inner product space and recall that $\mathcal{O}(V)$ is the group of orthogonal transformations on V to V.

Definition 2.7. A random vector X in $(V, (\cdot, \cdot))$ with distribution Q has an *orthogonally invariant distribution* if $\mathcal{L}(X) \equiv \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}(V)$, or equivalently if $Q(B) = Q(\Gamma B)$ for all Borel sets B and $\Gamma \in \mathcal{O}(V)$.

Many properties of orthogonally invariant distributions follow from the following proposition.

Proposition 2.10. Let $x_0 \in V$ with $\|x_0\| = 1$. If $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for $\Gamma \in \mathcal{O}(V)$, then for $x \in V$, $\mathcal{L}((x, X)) = \mathcal{L}(\|x\|(x_0, X))$.

Proof. The assertion is that the distribution of the real-valued random variable (x, X) is the same as the distribution of $\|x\|(x_0, X)$. Thus knowing the distribution of (x, X) for one particular nonzero $x \in V$ gives us the distribution of (x, X) for all $x \in V$. If $x = 0$, the assertion of the proposition is trivial. For $x \neq 0$, choose $\Gamma \in \mathcal{O}(V)$ such that $\Gamma x_0 = x/\|x\|$. This is possible since x_0 and $x/\|x\|$ both have norm 1. Thus

$$\begin{aligned}\mathcal{L}((x, X)) &= \mathcal{L}\left(\|x\|\left(\frac{x}{\|x\|}, X\right)\right) = \mathcal{L}\left(\|x\|(\Gamma x_0, X)\right) = \mathcal{L}\left(\|x\|(x_0, \Gamma' X)\right)\\ &= \mathcal{L}\left(\|x\|(x_0, X)\right)\end{aligned}$$

where the last equality follows from the assumption that $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}(V)$ and the fact that $\Gamma \in \mathcal{O}(V)$ implies $\Gamma' \in \mathcal{O}(V)$. □

Proposition 2.11. Let $x_0 \in V$ with $\|x_0\| = 1$. Suppose the distribution of X is orthogonally invariant. Then:

(i) $\phi(x) \equiv \mathcal{E} e^{i(x, X)} = \phi(\|x\| x_0)$.
(ii) If $\mathcal{E} X$ exists, then $\mathcal{E} X = 0$.
(iii) If $\text{Cov}(X)$ exists, then $\text{Cov}(X) = \sigma^2 I$ where $\sigma^2 = \text{var}\{(x_0, X)\}$, and I is the identity linear transformation.

Proof. Assertion (i) follows from Proposition 2.10 and

$$\mathcal{E}e^{i(x,X)} = \mathcal{E}e^{i\|x\|(x_0,X)} = \mathcal{E}e^{i(\|x\|x_0,X)} = \phi(\|x\|x_0).$$

For (ii), let $\mu = \mathcal{E}X$. Since $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$, $\mu = \mathcal{E}X = \mathcal{E}\Gamma X = \Gamma\mathcal{E}X = \Gamma\mu$ for all $\Gamma \in \mathcal{O}(V)$. The only vector μ that satisfies $\mu = \Gamma\mu$ for all $\Gamma \in \mathcal{O}(V)$ is $\mu = 0$. To prove (iii), we must show that $\sigma^2 I$ satisfies the defining equation for Cov(X). But by Proposition 2.10,

$$\text{var}\{(x, X)\} = \text{var}\{\|x\|(x_0, X)\} = \|x\|^2 \text{var}\{x_0, X\} = \sigma^2(x, x) = (x, \sigma^2 Ix)$$

so Cov(X) $= \sigma^2 I$ by Proposition 2.6. □

Assertion (i) of Proposition 2.11 shows that the characteristic function ϕ of an orthogonally invariant distribution satisfies $\phi(\Gamma x) = \phi(x)$ for all $x \in V$ and $\Gamma \in \mathcal{O}(V)$. Any function f defined on V and taking values in some set is called *orthogonally invariant* if $f(x) = f(\Gamma x)$ for all $\Gamma \in \mathcal{O}(V)$. A characterization of orthogonal invariant functions is given by the following proposition.

Proposition 2.12. A function f defined on $(V, (\cdot,\cdot))$ is orthogonally invariant iff $f(x) = f(\|x\|x_0)$ where $x_0 \in V$, $\|x_0\| = 1$.

Proof. If $f(x) = f(\|x\|x_0)$, then $f(\Gamma x) = f(\|\Gamma x\|x_0) = f(\|x\|x_0) = f(x)$ so f is orthogonally invariant. Conversely, suppose f is orthogonally invariant and $x_0 \in V$ with $\|x_0\| = 1$. For $x = 0$, $f(0) = f(\|x\|x_0)$ since $\|x\| = 0$. If $x \neq 0$, let $\Gamma \in \mathcal{O}(V)$ be such that $\Gamma x_0 = x/\|x\|$. Then $f(x) = f(\Gamma\|x\|x_0) = f(\|x\|x_0)$. □

If X has an orthogonally invariant distribution in $(V, (\cdot,\cdot))$ and h is a function on R to R, then

$$f(x) \equiv \mathcal{E}h((x, X))$$

clearly satisfies $f(\Gamma x) = f(x)$ for $\Gamma \in \mathcal{O}(V)$. Thus $f(x) = f(\|x\|x_0) = \mathcal{E}h(\|x\|(x_0, X))$, so to calculate $f(x)$, one only needs to calculate $f(\alpha x_0)$ for $\alpha \in (0, \infty)$. We have more to say about orthogonally invariant distributions in later chapters.

A random vector $X \in V(\cdot,\cdot)$ is called *orthogonally invariant* about x_0 if $X - x_0$ has an orthogonally invariant distribution. It is not difficult to show, using characteristic functions, that if X is orthogonally invariant about both x_0 and x_1, then $x_0 = x_1$. Further, if X is orthogonally invariant about x_0 and if $\mathcal{E}X$ exists, then $\mathcal{E}(X - x_0) = 0$ by Proposition 2.11. Thus $x_0 = \mathcal{E}X$ when $\mathcal{E}X$ exists.

It has been shown that if X has an orthogonally invariant distribution and if Cov(X) exists, then Cov(X) $= \sigma^2 I$ for some $\sigma^2 \geqslant 0$. Of course there are distributions other than orthogonally invariant distributions for which the covariance is a constant times the identity. Such distributions arise in the chapter on linear models.

Definition 2.8. If $X \in (V, (\cdot, \cdot))$ and

$$\operatorname{Cov}(X) = \sigma^2 I \quad \text{for some } \sigma^2 > 0,$$

X has a *weakly spherical* distribution.

The justification for the above definition is provided by Proposition 2.13.

Proposition 2.13. Suppose X is a random vector in $(V, (\cdot, \cdot))$ and Cov(X) exists. The following are equivalent:

(i) Cov(X) $= \sigma^2 I$ for some $\sigma^2 \geqslant 0$.
(ii) Cov(X) $=$ Cov(ΓX) for all $\Gamma \in \mathcal{O}(V)$.

Proof. That (i) implies (ii) follows from Proposition 2.8. To show (ii) implies (i), let $\Sigma = \operatorname{Cov}(X)$. From (ii) and Proposition 2.8, the non-negative definite linear transformation Σ must satisfy $\Sigma = \Gamma \Sigma \Gamma'$ for all $\Gamma \in \mathcal{O}(V)$. Thus for all $x \in V$, $\|x\| = 1$,

$$(x, \Sigma x) = (x, \Gamma \Sigma \Gamma' x) = (\Gamma' x, \Sigma \Gamma' x).$$

But $\Gamma' x$ can be any vector in V with length one since Γ' can be any element of $\mathcal{O}(V)$. Thus for all x, y, $\|x\| = \|y\| = 1$,

$$(x, \Sigma x) = (y, \Sigma y).$$

From the spectral theorem, write $\Sigma = \Sigma_1^n \lambda_i x_i \square x_i$ and choose $x = x_j$ and $y = x_k$. Then we have

$$\lambda_j = (x_j, \Sigma x_j) = (x_k, \Sigma x_k) = \lambda_k$$

for all j, k. Setting $\sigma^2 = \lambda_1$,

$$\Sigma = \Sigma_1^n \sigma^2 x_i \square x_i = \sigma^2 \Sigma_1^n x_i \square x_i = \sigma^2 I.$$

That $\sigma^2 \geqslant 0$ follows from the positive semidefiniteness of Σ. □

Orthogonally invariant distributions are sometimes called *spherical distributions*. The term weakly spherical results from weakening the assumption that the entire distribution is orthogonally invariant to the assumption that just the covariance structure is orthogonally invariant (condition (ii) of Proposition 2.13). A slight generalization of Proposition 2.13, given in its algebraic context, is needed for use later in this chapter.

Proposition 2.14. Suppose f is a bilinear function on $V \times V$ where $(V, (\cdot, \cdot))$ is an inner product space. If $f[\Gamma x_1, \Gamma x_2] = f[x_1, x_2]$ for all $x_1, x_2 \in V$ and $\Gamma \in \mathcal{O}(V)$, then $f[x_1, x_2] = c(x_1, x_2)$ where c is some real constant. If A is a linear transformation on V to V that satisfies $\Gamma' A \Gamma = A$ for all $\Gamma \in \mathcal{O}(V)$, then $A = cI$ for some real c.

Proof. Every bilinear function on $V \times V$ has the form (x_1, Ax_2) for some linear transformation A on V to V. The assertion that $f[\Gamma x_1, \Gamma x_2] = f[x_1, x_2]$ is clearly equivalent to the assertion that $\Gamma' A \Gamma = A$ for all $\Gamma \in \mathcal{O}(V)$. Thus it suffices to verify the assertion concerning the linear transformation A. Suppose $\Gamma' A \Gamma = A$ for all $\Gamma \in \mathcal{O}(V)$. Then for $x_1, x_2 \in V$,

$$(x_1, Ax_2) = (x_1, \Gamma' A \Gamma x_2) = (\Gamma x_1, A \Gamma x_2).$$

By Proposition 1.20, there exists a Γ such that

$$\Gamma \frac{x_1}{\|x_1\|} = \frac{x_2}{\|x_2\|}, \qquad \Gamma \frac{x_2}{\|x_2\|} = \frac{x_1}{\|x_1\|}$$

when x_1 and x_2 are not zero. Thus for x_1 and x_2 not zero,

$$(x_1, Ax_2) = (\Gamma x_1, A \Gamma x_2) = (x_2, Ax_1) = (Ax_1, x_2).$$

However, this relationship clearly holds if either x_1 or x_2 is zero. Thus for all $x_1, x_2 \in V$, $(x_1, Ax_2) = (Ax_1, x_2)$, so A must be self-adjoint. Now, using the spectral theorem, we can argue as in the proof of Proposition 2.13 to conclude that $A = cI$ for some real number c. □

◆ **Example 2.4.** Consider coordinate space R^n with the usual inner product. Let f be a function on $[0, \infty)$ to $[0, \infty)$ so that

$$\int_{R^n} f(\|x\|^2)\, dx = 1.$$

Thus $f(\|x\|^2)$ is a density on R^n. If the coordinate random vector

$X \in R^n$ has $f(\|x\|^2)$ as its density, then for $\Gamma \in \mathcal{O}_n$ (the group of $n \times n$ orthogonal matrices), the density of ΓX is again $f(\|x\|^2)$. This follows since $\|\Gamma x\| = \|x\|$ and the Jacobian of the linear transformation determined by Γ is equal to one. Hence the distribution determined by the density is $\mathcal{O}_n$ invariant. One particular choice for f is $f(u) = (2\pi)^{-n/2}e^{-1/2u}$ and the density for X is then

$$f(\|x\|^2) = (2\pi)^{-n/2}\exp\left[-\tfrac{1}{2}\Sigma_1^n x_i^2\right] = \prod_{i=1}^{n}(2\pi)^{-1/2}\exp\left[-\tfrac{1}{2}x_i^2\right].$$

Each of the factors in the above product is a density on R (corresponding to a normal distribution with mean zero and variance one). Therefore, the coordinates of X are independent and each has the same distribution. An example of a distribution on R^n that is weakly spherical, but not spherical, is provided by the density (with respect to Lebesgue measure)

$$p(x) = 2^{-n}\exp\left[-\Sigma_1^n|x_i|\right]$$

where $x \in R^n$, $x' = (x_1, x_2, \ldots, x_n)$. More generally, if the random variables $X_1, \ldots, X_n$ are independent with the same distribution on R, and $\sigma^2 = \text{var}(X_1)$, then the random vector X with coordinates $X_1, \ldots, X_n$ is easily shown to satisfy $\text{Cov}(X) = \sigma^2 I_n$ where I_n is the $n \times n$ identity matrix. ◆

The next topic in this section concerns the covariance between two random vectors. Suppose $X_i \in (V_i, (\cdot,\cdot)_i)$ for $i = 1, 2$ where X_1 and X_2 are defined on the same probability space. Then the random vector $\{X_1, X_2\}$ takes values in the direct sum $V_1 \oplus V_2$. Let $[\cdot,\cdot]$ denote the usual inner product on $V_1 \oplus V_2$ inherited from $(\cdot,\cdot)_i$, $i = 1, 2$. Assume that $\Sigma_{ii} = \text{Cov}(X_i)$, $i = 1, 2$, both exist. Then, let

$$f(x_1, x_2) = \text{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\}$$

and note that the Cauchy–Schwarz Inequality (Example 1.9) shows that

$$|f(x_1, x_2)|^2 \leqslant (x_1, \Sigma_{11}x_1)_1(x_2, \Sigma_{22}x_2)_2.$$

Further, it is routine to check that $f(\cdot,\cdot)$ is a bilinear function on $V_1 \times V_2$ so there exists a linear transformation $\Sigma_{12} \in \mathcal{L}(V_2, V_1)$ such that

$$f(x_1, x_2) = (x_1, \Sigma_{12}x_2)_1.$$

The next proposition relates Σ_{11}, Σ_{12}, and Σ_{22} to the covariance of $\{X_1, X_2\}$ in the vector space $(V_1 \oplus V_2, [\cdot, \cdot])$.

Proposition 2.15. Let $\Sigma = \text{Cov}\{X_1, X_2\}$. Define a linear transformation A on $V_1 \oplus V_2$ to $V_1 \oplus V_2$ by

$$A\{x_1, x_2\} = \{\Sigma_{11}x_1 + \Sigma_{12}x_2, \Sigma_{12}'x_1 + \Sigma_{22}x_2\}$$

where Σ_{12}' is the adjoint of Σ_{12}. Then $A = \Sigma$.

Proof. It is routine to check that

$$[A\{x_1, x_2\}, \{x_3, x_4\}] = [\{x_1, x_2\}, A\{x_3, x_4\}]$$

so A is self-adjoint. To show $A = \Sigma$, it is sufficient to verify

$$[\{x_1, x_2\}, A\{x_1, x_2\}] = [\{x_1, x_2\}, \Sigma\{x_1, x_2\}]$$

by Proposition 1.16. However,

$$\begin{aligned}
[\{x_1, x_2\}, \Sigma\{x_1, x_2\}] &= \text{var}[\{x_1, x_2\}, \{X_1, X_2\}] \\
&= \text{var}\{(x_1, X_1)_1 + (x_2, X_2)_2\} \\
&= \text{var}(x_1, X_1)_1 + \text{var}(x_2, X_2)_2 \\
&\quad + 2\,\text{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} \\
&= (x_1, \Sigma_{11}x_1)_1 + (x_2, \Sigma_{22}x_2)_2 + 2(x_1, \Sigma_{12}x_2)_1 \\
&= (x_1, \Sigma_{11}x_1)_1 + (x_2, \Sigma_{22}x_2)_2 \\
&\quad + (x_1, \Sigma_{12}x_2)_1 + (\Sigma_{12}'x_1, x_2)_2 \\
&= [\{x_1, x_2\}, \{\Sigma_{11}x_1 + \Sigma_{12}x_2, \Sigma_{12}'x_1 + \Sigma_{22}x_2\}] \\
&= [\{x_1, x_2\}, A\{x_1, x_2\}].
\end{aligned}$$ □

It is customary to write the linear transformation A in partitioned form as

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\{x_1, x_2\} = \{\Sigma_{11}x_1 + \Sigma_{12}x_2, \Sigma_{12}'x_1 + \Sigma_{22}x_2\}.$$

With this notation,

$$\mathrm{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}.$$

Definition 2.9. The random vectors X_1 and X_2 are uncorrelated if $\Sigma_{12} = 0$.

In the above definition, it is assumed that $\mathrm{Cov}(X_i)$ exists for $i = 1, 2$. It is clear that X_1 and X_2 are uncorrelated iff

$$\mathrm{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} = 0 \qquad \text{for all } x_i \in V_i,\ i = 1, 2.$$

Also, if X_1 and X_2 are uncorrelated in the two given inner products, then they are uncorrelated in all inner products on V_1 and V_2. This follows from the fact that any two inner products are related by a positive definite linear transformation.

Given $X_i \in (V_i, (\cdot, \cdot)_i)$ for $i = 1, 2$, suppose

$$\mathrm{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}.$$

We want to show that there is a linear transformation $B \in \mathcal{L}(V_2, V_1)$ such that $X_1 + BX_2$ and X_2 are uncorrelated random vectors. However, before this can be established, some preliminary technical results are needed.

Consider an inner product space $(V, (\cdot, \cdot))$ and suppose $A \in \mathcal{L}(V, V)$ is self-adjoint of rank k. Then, by the spectral theorem, $A = \Sigma_1^k \lambda_i x_i \Box x_i$ where $\lambda_i \neq 0$, $i = 1, \ldots, k$, and $\{x_1, \ldots, x_k\}$ is an orthonormal set that is a basis for $\mathcal{R}(A)$. The linear transformation

$$A^- \equiv \Sigma_1^k \frac{1}{\lambda_i} x_i \Box x_i$$

is called the *generalized inverse* of A. If A is nonsingular, then it is clear that A^- is the inverse of A. Also, A^- is self-adjoint and $AA^- = A^-A = \Sigma_1^k x_i \Box x_i$, which is just the orthogonal projection onto $\mathcal{R}(A)$. A routine computation shows that $A^-AA^- = A^-$ and $AA^-A = A$.

In the notation established previously (see Proposition 2.15), suppose $\{X_1, X_2\} \in V_1 \oplus V_2$ has a covariance

$$\Sigma = \mathrm{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}.$$

Proposition 2.16. For the covariance above, $\mathcal{N}(\Sigma_{22}) \subseteq \mathcal{N}(\Sigma_{12})$ and $\Sigma_{12} = \Sigma_{12}\Sigma_{22}^-\Sigma_{22}$.

Proof. For $x_2 \in \mathfrak{N}(\Sigma_{22})$, it must be shown that $\Sigma_{12}x_2 = 0$. Consider $x_1 \in V_1$ and $\alpha \in R$. Then $\Sigma_{22}(\alpha x_2) = 0$ and since Σ is positive semidefinite,

$$\begin{aligned} 0 &\leqslant [\{x_1, \alpha x_2\}, \Sigma\{x_1, \alpha x_2\}] = [\{x_1, \alpha x_2\}, \{\Sigma_{11}x_1 + \alpha\Sigma_{12}x_2, \Sigma'_{12}x_1\}] \\ &= (x_1, \Sigma_{11}x_1)_1 + \alpha(x_1, \Sigma_{12}x_2)_1 + \alpha(x_2, \Sigma'_{12}x_1)_2 \\ &= (x_1, \Sigma_{11}x_1) + 2\alpha(x_1, \Sigma_{12}x_2)_1. \end{aligned}$$

As this inequality holds for all $\alpha \in R$, for each $x_1 \in V$, $(x_1, \Sigma_{12}x_2)_1 = 0$. Hence $\Sigma_{12}x_2 = 0$ and the first claim is proved. To verify that $\Sigma_{12} = \Sigma_{12}\Sigma_{22}^-\Sigma_{22}$, it suffices to establish the identity $\Sigma_{12}(I - \Sigma_{22}^-\Sigma_{22}) = 0$. However, $I - \Sigma_{22}^-\Sigma_{22}$ is the orthogonal projection onto $\mathfrak{N}(\Sigma_{22})$. Since $\mathfrak{N}(\Sigma_{22}) \subseteq \mathfrak{N}(\Sigma_{12})$, it follows that $\Sigma_{12}(I - \Sigma_{22}^-\Sigma_{22}) = 0$. □

We are now in a position to show that $X_1 - \Sigma_{12}\Sigma_{22}^-X_2$ and X_2 are uncorrelated.

Proposition 2.17. Suppose $\{X_1, X_2\} \in V_1 \oplus V_2$ has a covariance

$$\Sigma = \text{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}.$$

Then $X_1 - \Sigma_{12}\Sigma_{22}^-X_2$ and X_2 are uncorrelated, and $\text{Cov}(X_1 - \Sigma_{12}\Sigma_{22}^-X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21}$ where $\Sigma_{21} \equiv \Sigma'_{12}$.

Proof. For $x_i \in V_i$, $i = 1, 2$, it must be verified that

$$\text{cov}\{(x_1, X_1 - \Sigma_{12}\Sigma_{22}^-X_2)_1, (x_2, X_2)_2\} = 0.$$

This calculation goes as follows:

$$\begin{aligned} &\text{cov}\{(x_1, X_1 - \Sigma_{12}\Sigma_{22}^-X_2)_1, (x_2, X_2)_2\} \\ &\quad = \text{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} \\ &\qquad - \text{cov}\{(\Sigma_{22}^-\Sigma'_{12}x_1, X_2)_2, (x_2, X_2)_2\} \\ &\quad = (x_1, \Sigma_{12}x_2)_1 - (\Sigma_{22}^-\Sigma'_{12}x_1, \Sigma_{22}x_2)_2 \\ &\quad = (x_1, \Sigma_{12}x_2)_1 - (x_1, \Sigma_{12}\Sigma_{22}^-\Sigma_{22}x_2)_1 \\ &\quad = (x_1, (\Sigma_{12} - \Sigma_{12}\Sigma_{22}^-\Sigma_{22})x_2)_1 = 0. \end{aligned}$$

The last equality follows from Proposition 2.15 since $\Sigma_{12} = \Sigma_{12}\Sigma_{22}^{-}\Sigma_{22}$. To verify the second assertion, we need to establish the identity

$$\operatorname{var}(x_1, X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2)_1 = (x_1, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21})x_1)_1.$$

But

$$\begin{aligned}\operatorname{var}(x_1, X_1 - \Sigma_{12}\Sigma_{22}^{-}X_2)_1 &= \operatorname{var}(x_1, X_1)_1 + \operatorname{var}(x_1, \Sigma_{12}\Sigma_{22}^{-}X_2)_1 \\ &\quad -2\operatorname{cov}\{(x_1, X_1)_1, (x_1, \Sigma_{12}\Sigma_{22}^{-}X_2)_1\} \\ &= (x_1, \Sigma_{11}x_1)_1 + (x_1, \Sigma_{12}\Sigma_{22}^{-}\Sigma_{22}\Sigma_{22}^{-}\Sigma_{12}'x_1)_1 \\ &\quad -2(x_1, \Sigma_{12}\Sigma_{22}^{-}\Sigma_{12}'x_1)_1 \\ &= (x_1, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{12}')x_1)_1.\end{aligned}$$

In the above, the identity $\Sigma_{22}^{-}\Sigma_{22}\Sigma_{22}^{-} = \Sigma_{22}^{-}$ has been used. □

We now return to the situation considered in Example 2.4. Consider independent coordinate random vectors $X_1, \ldots, X_n$ with each $X_i \in R^p$, and suppose that $\mathcal{E}X_i = \mu \in R^p$, and $\operatorname{Cov}(X_i) = \Sigma$ for $i = 1, \ldots, n$. Form the random matrix $X \in \mathcal{L}_{p,n}$ with rows $X_1', \ldots, X_n'$. Our purpose is to describe the mean vector and covariance of X in terms of Σ and μ. The inner product on $\mathcal{L}_{p,n}$, $\langle \cdot, \cdot \rangle$ is that inherited from the standard inner products on the coordinate spaces R^p and R^n. Recall that, for matrices $A, B \in \mathcal{L}_{p,n}$,

$$\langle A, B \rangle = \operatorname{tr} AB' = \operatorname{tr} B'A = \operatorname{tr} A'B = \operatorname{tr} BA'.$$

Let e denote the vector in R^n whose coordinates are all equal to 1.

Proposition 2.18. In the above notation,

(i) $\mathcal{E}X = e\mu'$.
(ii) $\operatorname{Cov}(X) = I_n \otimes \Sigma$.

Here I_n is the $n \times n$ identity matrix and $\otimes$ denotes the Kronecker product.

Proof. The matrix $e\mu'$ has each row equal to μ' and, since each row of X has mean μ', the first assertion is fairly obvious. To verify (i) formally, it must be shown that, for $A \in \mathcal{L}_{p,n}$,

$$\mathcal{E}\langle A, X \rangle = \langle A, e\mu' \rangle.$$

Let $a_1', \dots, a_n'$, $a_i \in R^p$, be the rows of A. Then

$$\mathcal{E}\langle A, X\rangle = \mathcal{E}\operatorname{tr} AX' = \mathcal{E}\Sigma_1^n a_i' X_i = \Sigma_1^n a_i' \mathcal{E} X_i = \Sigma_1^n a_i' \mu = \operatorname{tr} A\mu e' = \langle A, e\mu'\rangle.$$

Thus (i) holds. To verify (ii) it suffices to establish the identity

$$\operatorname{var}\langle A, X\rangle = \langle A, (I \otimes \Sigma)A\rangle$$

for $A \in \mathcal{L}_{p,n}$. In the notation above,

$$\operatorname{var}\langle A, X\rangle = \operatorname{var}(\Sigma_1^n a_i' X_i) = \Sigma_1^n \operatorname{var}(a_i' X_i) + \sum_{i \neq j} \operatorname{cov}\{a_i' X_i, a_j' X_j\} = \Sigma_1^n a_i' \Sigma a_i$$

$$= \operatorname{tr} A' A \Sigma = \operatorname{tr} A \Sigma A' = \operatorname{tr} A(A\Sigma)' = \langle A, (I_n \otimes \Sigma)A\rangle.$$

The third equality follows from $\operatorname{var}(a_i' X) = a_i' \Sigma a_i$ and, for $i \neq j$, $a_i' X_i$ and $a_j' X_j$ are uncorrelated. □

The assumption of the independence of $X_1, \dots, X_n$ was not used to its full extent in the proof of Proposition 2.18. In fact the above proof shows that, if $X_1, \dots, X_n$ are random variables in R^p with $\mathcal{E} X_i = \mu$, $i = 1, \dots, n$, then $\mathcal{E} X = e\mu'$. Further, if $X_1, \dots, X_n$ in R^p are uncorrelated with $\operatorname{Cov}(X_i) = \Sigma$, $i = 1, \dots, n$, then $\operatorname{Cov}(X) = I_n \otimes \Sigma$. One application of this formula for $\operatorname{Cov}(X)$ describes how $\operatorname{Cov}(X)$ transforms under Kronecker products. For example, if $A \in \mathcal{L}_{n,n}$ and $B \in \mathcal{L}_{p,p}$, then $(A \otimes B)X = AXB'$ is a random vector in $\mathcal{L}_{p,n}$. Proposition 2.8 shows that

$$\operatorname{Cov}((A \otimes B)X) = (A \otimes B)\operatorname{Cov}(X)(A \otimes B)'.$$

In particular, if $\operatorname{Cov}(X) = I_n \otimes \Sigma$, then

$$\operatorname{Cov}((A \otimes B)X) = (A \otimes B)(I_n \otimes \Sigma)(A \otimes B)' = (AA') \otimes (B\Sigma B').$$

Since $A \otimes B = (A \otimes I_p)(I_n \otimes B)$, the interpretation of the above covariance formula reduces to an interpretation for $A \otimes I_p$ and $I_n \otimes B$. First, $(I_n \otimes B)X$ is a random matrix with rows $X_i'B' = (BX_i)'$, $i = 1, \dots, n$. If $\operatorname{Cov}(X_i) = \Sigma$, then $\operatorname{Cov}(BX_i) = B\Sigma B'$. Thus it is clear from Proposition 2.18 that $\operatorname{Cov}((I_n \times B)X) = I_n \otimes (B\Sigma B')$. Second, $(A \otimes I_p)$ applied to X is the same as applying the linear transformation A to each column of X. When $\operatorname{Cov}(X) = I_n \otimes \Sigma$, the rows of X are uncorrelated and, if A is an $n \times n$ orthogonal matrix, then

$$\operatorname{Cov}\big((A \otimes I_p)X\big) = I_n \otimes \Sigma = \operatorname{Cov}(X).$$

Thus the absence of correlation between the rows is preserved by an orthogonal transformation of the columns of X.

A converse to the observation that $\text{Cov}((A \otimes I_p)X) = I_n \otimes \Sigma$ for all $A \in \mathcal{O}(n)$ is valid for random linear transformations. To be more precise, we have the following proposition.

Proposition 2.19. Suppose $(V_i, (\cdot, \cdot)_i)$, $i = 1, 2$, are inner product spaces and X is a random vector in $(\mathcal{L}(V_1, V_2), \langle \cdot , \cdot \rangle)$. The following are equivalent:

(i) $\text{Cov}(X) = I_2 \otimes \Sigma$.
(ii) $\text{Cov}((\Gamma \otimes I_1)X) = \text{Cov}(X)$ for all $\Gamma \in \mathcal{O}(V_2)$.

Here, I_i is identity linear transformation on V_i, $i = 1, 2$, and Σ is a non-negative definite linear transformation on V_1 to V_1.

Proof. Let $\Psi = \text{Cov}(X)$ so Ψ is a positive semidefinite linear transformation on $\mathcal{L}(V_1, V_2)$ to $\mathcal{L}(V_1, V_2)$ and Ψ is characterized by the equation

$$\text{cov}\{\langle A, X \rangle, \langle B, X \rangle\} = \langle A, \Psi B \rangle$$

for all $A, B \in \mathcal{L}(V_1, V_2)$. If (i) holds, then we have

$$\begin{aligned}\text{Cov}((\Gamma \otimes I_1)X) &= (\Gamma \otimes I_1)\text{Cov}(X)(\Gamma \otimes I_1)' \\ &= (\Gamma \otimes I_1)(I_2 \otimes \Sigma)(\Gamma' \otimes I_1) = (\Gamma I_2 \Gamma') \otimes (I_1 \Sigma I_1) \\ &= I_2 \otimes \Sigma = \text{Cov}(X),\end{aligned}$$

so (ii) holds.

Now, assume (ii) holds. Since outer products form a basis for $\mathcal{L}(V_1, V_2)$, it is sufficient to show there exists a positive semidefinite Σ on V_1 to V_1 such that, for $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$,

$$\langle y_1 \square x_1, \Psi(y_2 \square x_2) \rangle = \langle y_1 \square x_1, (I_2 \otimes \Sigma)(y_2 \square x_2) \rangle.$$

Define H by

$$H(x_1, x_2, y_1, y_2) \equiv \text{cov}\{\langle y_1 \square x_1, X \rangle, \langle y_2 \square x_2, X \rangle\}$$

for $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$. From assumption (ii), we know that Ψ

satisfies $\Psi = (\Gamma \otimes I_1)\Psi(\Gamma \otimes I_1)'$ for all $\Gamma \in \mathcal{O}(V_2)$. Thus

$$\begin{aligned} H(x_1, x_2, y_1, y_2) &= \langle y_1 \square x_1, \Psi(y_2 \square x_2)\rangle \\ &= \langle y_1 \square x_1, (\Gamma \otimes I_1)\Psi(\Gamma \otimes I_1)'(y_2 \square x_2)\rangle \\ &= \langle (\Gamma \otimes I_1)'(y_1 \square x_1), \Psi(\Gamma \otimes I_1)'(y_2 \square x_2)\rangle \\ &= \langle (\Gamma' y_1) \square x_1, \Psi(\Gamma' y_2) \square x_2\rangle = H(x_1, x_2, \Gamma' y_1, \Gamma' y_2) \end{aligned}$$

for all $\Gamma \in \mathcal{O}(V_2)$. It is clear that H is a linear function of each of its four arguments when the other three are held fixed. Therefore, for x_1 and x_2 fixed, G is a bilinear function on $V_2 \times V_2$ and this bilinear function satisfies the assumption of Proposition 2.14. Thus there is a constant, which depends on x_1 and x_2, say $c[x_1, x_2]$, and

$$H(x_1, x_2, y_1, y_2) = c[x_1, x_2](y_1, y_2)_2.$$

However, for $y_1 = y_2 \neq 0$, H, as a function of x_1 and x_2, is bilinear and non-negative definite on $V_1 \times V_1$. In other words, $c[x_1, x_2]$ is a non-negative definite bilinear function on $V_1 \times V_1$, so

$$c[x_1, x_2] = (x_1, \Sigma x_2)_1$$

for some non-negative definite Σ. Thus

$$H(x_1, x_2, y_1, y_2) = (x_1, \Sigma x_2)_1 (y_1, y_2)_2 = \langle y_1 \square x_1, (I_2 \otimes \Sigma)(y_2 \square x_2)\rangle,$$

so $\Psi = I_2 \otimes \Sigma$. □

The next topic of consideration in the section concerns the calculation of means and covariances for outer products of random vectors. These results are used throughout the sequel to simplify proofs and provide convenient formulas. Suppose X_i is a random vector in $(V_i, (\cdot, \cdot)_i)$ for $i = 1, 2$ and let $\mu_i = \mathcal{E} X_i$, and $\Sigma_{ii} = \text{Cov}(X_i)$ for $i = 1, 2$. Thus $\{X_1, X_2\}$ takes values in $V_1 \oplus V_2$ and

$$\text{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$$

where Σ_{12} is characterized by

$$\text{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} = (x_1, \Sigma_{12} x_2)_1$$

for $x_i \in V_i$, $i = 1, 2$. Of course, $\text{Cov}\{X_1, X_2\}$ is expressed relative to the natural inner product on $V_1 \oplus V_2$ inherited from $(V_1, (\cdot, \cdot)_1)$ and $(V_2, (\cdot, \cdot)_2)$.

Proposition 2.20. For $X_i \in (V_i, (\cdot, \cdot))$, $i = 1, 2$, as above,

$$\mathcal{E} X_1 \,\Box\, X_2 = \Sigma_{12} + \mu_1 \,\Box\, \mu_2.$$

Proof. The random vector $X_1 \,\Box\, X_2$ takes values in the inner product space $(\mathcal{L}(V_2, V_1), \langle \cdot, \cdot \rangle)$. To verify the above formula, it must be shown that

$$\mathcal{E}\langle A, X_1 \,\Box\, X_2 \rangle = \langle A, \Sigma_{12} \rangle + \langle A, \mu_1 \,\Box\, \mu_2 \rangle$$

for $A \in \mathcal{L}(V_2, V_1)$. However, it is sufficient to verify this equation for $A = x_1 \,\Box\, x_2$ since both sides of the equation are linear in A and every A is a linear combination of elements in $\mathcal{L}(V_2, V_1)$ of the form $x_1 \,\Box\, x_2$, $x_i \in V_i$, $i = 1, 2$. For $x_1 \,\Box\, x_2 \in \mathcal{L}(V_2, V_1)$,

$$\begin{aligned} \mathcal{E}\langle x_1 \,\Box\, x_2, X_1 \,\Box\, X_2 \rangle &= \mathcal{E}(x_1, X_1)_1 (x_2, X_2)_2 \\ &= \text{cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} + \mathcal{E}(x_1, X_1)_1 \mathcal{E}(x_2, X_2)_2 \\ &= (x_1, \Sigma_{12} x_2)_1 + (x_1, \mu_1)_1 (x_2, \mu_2)_2 \\ &= \langle x_1 \,\Box\, x_2, \Sigma_{12} \rangle + \langle x_1 \,\Box\, x_2, \mu_1 \,\Box\, \mu_2 \rangle. \end{aligned}$$

$\Box$

A couple of interesting applications of Proposition 2.20 are given in the following proposition.

Proposition 2.21. For X_1, X_2 in $(V, (\cdot, \cdot))$, let $\mu_i = \mathcal{E} X_i$, $\Sigma_{ii} = \text{Cov}(X_i)$ for $i = 1, 2$. Also, let Σ_{12} be the unique linear transformation satisfying

$$\text{cov}\{(x_1, X_1), (x_2, X_2)\} = (x_1, \Sigma_{12} x_2)$$

for all $x_1, x_2 \in V$. Then:

(i) $\mathcal{E} X_1 \,\Box\, X_1 = \Sigma_{11} + \mu_1 \,\Box\, \mu_1$.
(ii) $\mathcal{E}(X_1, X_2) = \langle I, \Sigma_{12} \rangle + (\mu_1, \mu_2)$.
(iii) $\mathcal{E}(X_1, X_1) = \langle I, \Sigma_{11} \rangle + (\mu_1, \mu_1)$.

Here $I \in \mathcal{L}(V, V)$ is the identity linear transformation and $\langle \cdot, \cdot \rangle$ is the inner product on $\mathcal{L}(V, V)$ inherited from $(V, (\cdot, \cdot))$.

Proof. For (i), take $X_1 = X_2$ and $(V_1, (\cdot, \cdot)_1) = (V_2, (\cdot, \cdot)_2) = (V, (\cdot, \cdot))$ in Proposition 2.20. To verify (ii), first note that

$$\mathcal{E} X_1 \square X_2 = \Sigma_{12} + \mu_1 \square \mu_2$$

by the previous proposition. Thus for $I \in \mathcal{L}(V, V)$,

$$\mathcal{E}\langle I, X_1 \square X_2 \rangle = \langle I, \Sigma_{12} \rangle + \langle I, \mu_1 \square \mu_2 \rangle.$$

However, $\langle I, X_1 \square X_2 \rangle = (X_1, X_2)$ and $\langle I, \mu_1 \square \mu_2 \rangle = (\mu_1, \mu_2)$ so (ii) holds. Assertion (iii) follows from (ii) by taking $X_1 = X_2$. □

One application of the preceding result concerns the affine prediction of one random vector by another random vector. By an affine function on a vector space V to W, we mean a function f given by $f(v) = Av + w_0$ where $A \in \mathcal{L}(V, W)$ and w_0 is a fixed vector in W. The term linear transformation is reserved for those affine functions that map zero into zero. In the notation of Proposition 2.21, consider $X_i \in (V_i, (\cdot, \cdot)_i$ for $i = 1, 2$, let $\mu_i = \mathcal{E} X_i$, $i = 1, 2$, and suppose

$$\Sigma \equiv \operatorname{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}$$

exists. An affine predictor of X_2 based on X_1 is any function of the form $AX_1 + x_0$ where $A \in \mathcal{L}(V_1, V_2)$ and x_0 is a fixed vector in V_2. If we assume that μ_1, μ_2, and Σ are known, then A and x_0 are allowed to depend on these known quantities. The statistical interpretation is that we observe X_1, but not X_2, and X_2 is to be predicted by $AX_1 + x_0$. One intuitively reasonable criterion for selecting A and x_0 is to ask that the choice of A and x_0 minimize

$$\mathcal{E}\| X_2 - (AX_1 + x_0) \|_2^2.$$

Here, the expectation is over the joint distribution of X_1 and X_2 and $\| \cdot \|_2$ is the norm in the vector space $(V_2, (\cdot, \cdot)_2)$. The quantity $\mathcal{E}\| X_2 - (AX_1 + x_0) \|_2^2$ is the average distance of $X_2 - (AX_1 + x_0)$ from 0. Since $AX_1 + x_0$ is supposed to predict X_2, it is reasonable that A and x_0 be chosen to minimize this average distance. A solution to this minimization problem is given in Proposition 2.22.

Proposition 2.22. For X_1 and X_2 as above,

$$\mathcal{E}\| X_2 - (AX_1 + x_0) \|_2^2 \geqslant \langle I_2, \Sigma_{22} - \Sigma'_{12} \Sigma_{11}^{-} \Sigma_{12} \rangle$$

with equality for $A = \Sigma'_{12} \Sigma_{11}^{-}$ and $x_0 = \mu_2 - \Sigma'_{12} \Sigma_{11}^{-} \mu_1$.

Proof. The proof is a calculation. It essentially consists of completing the square and applying (ii) of Proposition 2.21. Let $Y_i = X_i - \mu_i$ for $i = 1, 2$. Then

$$\mathcal{E}\|X_2 - (AX_1 + x_0)\|_2^2 = \mathcal{E}\|Y_2 - AY_1 + \mu_2 - A\mu_1 - x_0\|_2^2 = \mathcal{E}\|Y_2 - AY_1\|_2^2$$

$$+2\mathcal{E}(Y_2 - AY_1, \mu_2 - A\mu_1 - x_0)_2 + \|\mu_2 - A\mu_1 - x_0\|_2^2$$

$$= \mathcal{E}\|Y_2 - AY_1\|_2^2 + \|\mu_2 - A\mu_1 - x_0\|_2^2.$$

The last equality holds since $\mathcal{E}(Y_2 - AY_1) = 0$. Thus for each $A \in \mathcal{L}(V_1, V_2)$,

$$\mathcal{E}\|X_2 - (AX_1 + x_0)\|_2^2 \geqslant \mathcal{E}\|Y_2 - AY_1\|_2^2$$

with equality for $x_0 = \mu_2 - A\mu_1$. For notational convenience let $\Sigma_{21} = \Sigma_{12}'$. Then

$$\mathcal{E}\|Y_2 - AY_1\|_2^2 = \mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1 + (\Sigma_{21}\Sigma_{11}^- - A)Y_1\|_2^2$$

$$= \mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1\|_2^2 + \mathcal{E}\|(\Sigma_{21}\Sigma_{11}^- - A)Y_1\|_2^2$$

$$+2\mathcal{E}(Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1, (\Sigma_{21}\Sigma_{11}^- - A)Y_1)_2$$

$$= \mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1\|_2^2 + \mathcal{E}\|(\Sigma_{21}\Sigma_{11}^- - A)Y_1\|_2^2$$

$$\geqslant \mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1\|_2^2.$$

The last equality holds since $\mathcal{E}(Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1) = 0$ and $Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1$ is uncorrelated with Y_1 (Proposition 2.17) and hence is uncorrelated with $(\Sigma_{21}\Sigma_{11}^- - A)Y_1$. By (ii) of Proposition 2.21, we see that $\mathcal{E}(Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1, (\Sigma_{21}\Sigma_{11}^- - A)Y_1)_2 = 0$. Therefore, for each $A \in \mathcal{L}(V_1, V_2)$,

$$\mathcal{E}\|Y_2 - AY_1\|_2^2 \geqslant \mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1\|_2^2$$

with equality for $A = \Sigma_{21}\Sigma_{11}^-$. However, $\text{Cov}(Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^-\Sigma_{12}$ and $\mathcal{E}(Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1) = 0$ so (iii) of Proposition 2.21 shows that

$$\mathcal{E}\|Y_2 - \Sigma_{21}\Sigma_{11}^- Y_1\|_2^2 = \langle I_2, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^-\Sigma_{12}\rangle.$$

Therefore,

$$\mathcal{E}\|X_2 - (AX_1 + x_0)\|_2^2 \geqslant \langle I_2, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^-\Sigma_{12}\rangle$$

with equality for $A = \Sigma_{21}\Sigma_{11}^-$ and $x_0 = \mu_2 - \Sigma_{21}\Sigma_{11}^-\mu_1$. □

The last topic in this section concerns the covariance of $X \Box X$ when X is a random vector in $(V, (\cdot, \cdot))$. The random vector $X \Box X$ is an element of the vector space $(\mathcal{L}(V, V), \langle \cdot, \cdot \rangle)$. However, $X \Box X$ is a self-adjoint linear transformation so $X \Box X$ is also a random vector in $(M_s, \langle \cdot, \cdot \rangle)$ where M_s is the linear subspace of self-adjoint transformations in $\mathcal{L}(V, V)$. In what follows, we regard $X \Box X$ as a random vector in $(M_s, \langle \cdot, \cdot \rangle)$. Thus the covariance of $X \Box X$ is a positive semidefinite linear transformation on $(M_s, \langle \cdot, \cdot \rangle)$. In general, this covariance is quite complicated and we make some simplifying assumptions concerning the distribution of X.

Proposition 2.23. Suppose X has an orthogonally invariant distribution in $(V, (\cdot, \cdot))$ where $\mathcal{E}\|X\|^4 < +\infty$. Let v_1 and v_2 be fixed vectors in V with $\|v_i\| = 1$, $i = 1, 2$, and $(v_1, v_2) = 0$. Set $c_1 = \operatorname{var}\{(v_1, X)^2\}$ and $c_2 = \operatorname{cov}\{(v_1, X)^2, (v_2, X)^2\}$. Then

$$\operatorname{Cov}(X \Box X) = (c_1 - c_2) I \otimes I + c_2 T_1,$$

where T_1 is the linear transformation on M_s given by $T_1(A) = \langle I, A \rangle I$. In other words, for $A, B \in M_s$,

$$\begin{aligned}\operatorname{cov}\{\langle A, X \Box X \rangle, \langle B, X \Box X \rangle\} &= \langle A, ((c_1 - c_2) I \otimes I + c_2 T_1) B \rangle \\ &= (c_1 - c_2)\langle A, B \rangle + c_2 \langle I, A \rangle \langle I, B \rangle.\end{aligned}$$

Proof. Since $(c_1 - c_2) I \otimes I + c_2 T_1$ is self-adjoint on $(M_s, \langle \cdot, \cdot \rangle)$, Proposition 2.6 shows that it suffices to verify the equation

$$\operatorname{var}\langle A, X \Box X \rangle = (c_1 - c_2)\langle A, A \rangle + c_2 \langle I, A \rangle^2$$

for $A \in M_s$ in order to prove that

$$\operatorname{Cov}(X \Box X) = (c_1 - c_2) I \otimes I + c_2 T_1.$$

First note that, for $x \in V$,

$$\operatorname{var}\langle x \Box x, X \Box X \rangle = \operatorname{var}(x, X)^2 = \|x\|^4 \operatorname{var}\left(\frac{x}{\|x\|}, X\right)^2 = \|x\|^4 \operatorname{var}(v_1, X)^2.$$

This last equality follows from Proposition 2.10 as the distribution of X is orthogonally invariant. Also, for $x_1, x_2 \in V$ with $(x_1, x_2) = 0$,

$$\begin{aligned}\operatorname{cov}\{(x_1, X)^2, (x_2, X)^2\} &= \|x_1\|^2\|x_2\|^2 \operatorname{cov}\left\{\left(\frac{x_1}{\|x_1\|}, X\right)^2, \left(\frac{x_2}{\|x_2\|}, X\right)^2\right\} \\ &= \|x_1\|^2\|x_2\|^2 \operatorname{cov}\{(v_1, X)^2, (v_2, X)^2\}.\end{aligned}$$

Again, the last equality follows since $\mathcal{L}(X) = \mathcal{L}(\Psi X)$ for $\Psi \in \mathcal{O}(V)$ so

$$\operatorname{cov}\left\{\left(\frac{x_1}{\|x_1\|}, X\right)^2, \left(\frac{x_2}{\|x_2\|}, X\right)^2\right\} = \operatorname{cov}\left\{\left(\Psi\frac{x_1}{\|x_1\|}, X\right)^2, \left(\Psi\frac{x_2}{\|x_2\|}, X\right)^2\right\}$$

and Ψ can be chosen so that

$$\Psi\frac{x_i}{\|x_i\|} = v_i, \qquad i = 1, 2.$$

For $A \in M_s$, apply the spectral theorem and write $A = \sum_1^n a_i x_i \Box x_i$ where $x_1, \ldots, x_n$ is an orthonormal basis for $(V, (\cdot, \cdot))$. Then

$$\begin{aligned}\operatorname{var}\langle A, X \Box X\rangle &= \operatorname{var}\langle \Sigma a_i x_i \Box x_i, X \Box X\rangle \\ &= \Sigma a_i^2 \operatorname{var}\langle x_i \Box x_i, X \Box X\rangle \\ &\quad + \sum_{i \neq j}\sum a_i a_j \operatorname{cov}\{\langle x_i \Box x_i, X \Box X\rangle, \langle x_j \Box x_j, X \Box X\rangle\} \\ &= \Sigma a_i^2 \operatorname{var}(x_i, X)^2 + \sum_{i \neq j}\sum a_i a_j \operatorname{cov}\{(x_i, X)^2, (x_j, X)^2\} \\ &= c_1 \Sigma a_i^2 + c_2 \sum_{i \neq j}\sum a_i a_j = (c_1 - c_2)\sum_i a_i^2 + c_2 \sum_i \sum_j a_i a_j \\ &= (c_1 - c_2)\langle A, A\rangle + c_2 \langle I, A\rangle^2. \qquad \Box\end{aligned}$$

When X has an orthogonally invariant normal distribution, then the constant $c_2 = 0$ so $\operatorname{Cov}(X \Box X) = c_1 I \otimes I$. The following result provides a slight generalization of Proposition 2.23.

Proposition 2.24. Let X, v_1, and v_2 be as in Proposition 2.23. For $C \in \mathcal{L}(V, V)$, let $\Sigma = CC'$ and suppose Y is a random vector in $(V, (\cdot, \cdot))$ with

$\mathcal{L}(Y) = \mathcal{L}(CX)$. Then

$$\text{Cov}(Y \square Y) = (c_1 - c_2)\Sigma \otimes \Sigma + c_2 T_2$$

where $T_2(A) = \langle A, \Sigma \rangle \Sigma$ for $A \in M_s$.

Proof. We apply Proposition 2.8 and the calculational rules for Kronecker products. Since $(CX)\square(CX) = (C \otimes C)(X \square X)$,

$$\begin{aligned}
\text{Cov}(Y \square Y) &= \text{Cov}((CX \square CX)) = \text{Cov}((C \otimes C)(X \square X)) \\
&= (C \otimes C)\text{Cov}(X \square X)(C \otimes C)' \\
&= (C \otimes C)((c_1 - c_2)I \otimes I + c_2 T_1)(C' \otimes C') \\
&= (c_1 - c_2)(C \otimes C)(I \otimes I)(C' \otimes C') \\
&\quad + c_2(C \otimes C)T_1(C' \otimes C') \\
&= (c_1 - c_2)\Sigma \otimes \Sigma + c_2(C \otimes C)T_1(C' \otimes C').
\end{aligned}$$

It remains to show that $(C \otimes C)T_1(C' \otimes C') = T_2$. For $A \in M_s$,

$$\begin{aligned}
(C \otimes C)T_1(C' \otimes C')(A) &= C \otimes C(\langle I, (C' \otimes C')A \rangle I) \\
&= \langle (C \otimes C)I, A \rangle (C \otimes C)(I) = \langle CC', A \rangle CC' \\
&= \langle \Sigma, A \rangle \Sigma = T_2(A).
\end{aligned}$$

□

PROBLEMS

1. If $x_1, \ldots, x_n$ is a basis for $(V, (\cdot, \cdot))$ and if (x_i, X) has finite expectation for $i = 1, \ldots, n$, show that (x, X) has finite expectation for all $x \in V$. Also, show that if $(x_i, X)^2$ has finite expectation for $i = 1, \ldots, n$, then $\text{Cov}(X)$ exists.

2. Verify the claim that if $X_1(X_2)$ with values in $V_1(V_2)$ are uncorrelated for one pair of inner products on V_1 and V_2, then they are uncorrelated no matter what the inner products are on V_1 and V_2.

3. Suppose $X_i \in V_i$, $i = 1, 2$ are uncorrelated. If f_i is a linear function on V_i, $i = 1, 2$, show that

   (2.2) $$\text{cov}\{f_1(X_1), f_2(X_2)\} = 0.$$

   Conversely, if (2.2) holds for all linear functions f_1 and f_2, then X_1 and X_2 are uncorrrelated (assuming the relevant expectations exist).

4. For $X \in R^n$, partition X as

$$X = \begin{pmatrix} \dot{X} \\ \ddot{X} \end{pmatrix}$$

with $\dot{X} \in R^r$ and suppose X has an orthogonally invariant distribution. Show that $\dot{X}$ has an orthogonally invariant distribution on R^r. Argue that the conditional distribution of $\dot{X}$ given $\ddot{X}$ has an orthogonally invariant distribution.

5. Suppose $X_1, \dots, X_k$ in $(V, (\cdot, \cdot))$ are pairwise uncorrelated. Prove that $\text{Cov}(\Sigma_1^k X_i) = \Sigma_1^k \text{Cov}(X_i)$.

6. In R^k, let $\varepsilon_1, \dots, \varepsilon_k$ denote the standard basis vectors. Define a random vector U in R^k by specifying that U takes on the value ε_i with probability p_i where $0 \leqslant p_i \leqslant 1$ and $\Sigma_1^k p_i = 1$. (U represents one of k mutually exclusive and exhaustive events that can occur). Let $p \in R^k$ have coordinates $p_1, \dots, p_k$. Show that $\mathcal{E}U = p$, $\text{Cov}(U) = D_p - pp'$ where D_p is a diagonal matrix with diagonal entries $p_1, \dots, p_k$. When $0 < p_i < 1$, show that $\text{Cov}(U)$ has rank $k - 1$ and identify the null space of $\text{Cov}(U)$. Now, let $X_1, \dots, X_n$ be i.i.d. each with the distribution of U. The random vector $Y = \Sigma_1^n X_i$ has a multinomial distribution (prove this) with parameters k (the number of cells), the vector of probabilities p, and the number of trials n. Show that $\mathcal{E}Y = np$, $\text{Cov}(Y) = n(D_p - pp')$.

7. Fix a vector x in R^n and let π denote a permutation of $1, 2, \dots, n$ (there are $n!$ such permutations). Define the permuted vector πx to be the vector whose ith coordinate is $x(\pi^{-1}(i))$ where $x(j)$ denotes the jth coordinate of x. (This choice is justified in Chapter 7.) Let X be a random vector such that $P_r\{X = \pi x\} = 1/n!$ for each possible permutation π. Find $\mathcal{E}X$ and $\text{Cov}(X)$.

8. Consider a random vector $X \in R^n$ and suppose $\mathcal{L}(X) = \mathcal{L}(DX)$ for each diagonal matrix D with diagonal elements $d_{ii} = \pm 1$, $i = 1, \dots, n$. If $\mathcal{E}\|X\|^2 < +\infty$, show that $\mathcal{E}X = 0$ and $\text{Cov}(X)$ is a diagonal matrix (the coordinates of X are uncorrelated).

9. Given $X \in (V, (\cdot, \cdot))$ with $\text{Cov}(X) = \Sigma$, let A_i be a linear transformation on $(V, (\cdot, \cdot))$ to $(W_i, [\cdot, \cdot]_i)$, $i = 1, 2$. Form $Y = \{A_1 X, A_2 X\}$ with values in the direct sum $W_1 \oplus W_2$. Show

$$\text{Cov}(Y) = \begin{pmatrix} A_1 \Sigma A_1' & A_1 \Sigma A_2' \\ A_2 \Sigma A_1' & A_2 \Sigma A_2' \end{pmatrix}$$

in $W_1 \oplus W_2$ with its usual inner product.

10. For X in $(V, (\cdot, \cdot))$ with $\mu = \mathcal{E}X$ and $\Sigma = \text{Cov}(X)$, show that $\mathcal{E}(X, AX) = \langle A, \Sigma \rangle + (\mu, A\mu)$ for any $A \in \mathcal{L}(V, V)$.

11. In $(\mathcal{L}_{p,n}, \langle \cdot , \cdot \rangle)$, suppose the $n \times p$ random matrix X has the covariance $I_n \otimes \Sigma$ for some $p \times p$ positive semidefinite Σ. Show that the rows of X are uncorrelated. If $\mu = \mathcal{E}X$ and A is an $n \times n$ matrix, show that $\mathcal{E}X'AX = (\text{tr}\, A)\Sigma + \mu' A \mu$.

12. The usual inner product on the space of $p \times p$ symmetric matrices, denoted by $\mathcal{S}_p$, is $\langle \cdot , \cdot \rangle$, given by $\langle A, B \rangle = \text{tr}\, AB'$. (This is the natural inner product inherited from $(\mathcal{L}_{p,p}, \langle \cdot , \cdot \rangle)$ by regarding $\mathcal{S}_p$ as a subspace of $\mathcal{L}_{p,p}$.) Let S be a random matrix with values in $\mathcal{S}_p$ and suppose that $\mathcal{L}(\Gamma S \Gamma') = \mathcal{L}(S)$ for all $\Gamma \in \mathcal{O}_p$. (For example, if $X \in R^p$ has an orthogonally invariant distribution and $S = XX'$, then $\mathcal{L}(\Gamma S \Gamma') = \mathcal{L}(S)$.) Show that $\mathcal{E}S = cI_p$ where c is constant.

13. Given a random vector X in $(\mathcal{L}(V, W), \langle \cdot , \cdot \rangle)$, suppose that $\mathcal{L}(X) = \mathcal{L}((\Gamma \otimes \psi)X)$ for all $\Gamma \in \mathcal{O}(W)$ and $\psi \in \mathcal{O}(V)$.

(i) If X has a covariance, show $\mathcal{E}X = 0$ and $\text{Cov}(X) = cI_W \otimes I_V$ where $c \geqslant 0$.

(ii) If $Y \in \mathcal{L}(V, W)$ has a density (with respect to Lebesgue measure) given by $f(y) = p(\langle y, y \rangle)$, $y \in \mathcal{L}(V, W)$, show that $\mathcal{L}(Y) = \mathcal{L}((\Gamma \otimes \psi)Y)$ for $\Gamma \in \mathcal{O}(W)$ and $\psi \in \mathcal{O}(V)$.

14. Let $X_1, \ldots, X_n$ be uncorrelated random vectors in R^p with $\text{Cov}(X_i) = \Sigma$, $i = 1, \ldots, n$. Form the $n \times p$ random matrix X with rows $X_1', \ldots, X_n'$ and values in $(\mathcal{L}_{p,n}, \langle \cdot , \cdot \rangle)$. Thus $\text{Cov}(X) = I_n \otimes \Sigma$.

(i) Form $\tilde{X}$ in the coordinate space R^{np} with the coordinate inner product where

$$\tilde{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

In the space R^{np} show that

$$\text{Cov}(\tilde{X}) = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma \end{pmatrix}$$

where each block is $p \times p$.

(ii) Now, form $\tilde{\tilde{X}}$ in the space R^{np} where

$$\tilde{\tilde{X}} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}; \qquad Z_i \in R^n$$

and Z_i has coordinates $X_{1i}, \ldots, X_{ni}$ for $i = 1, \ldots, p$. Show that

$$\operatorname{Cov}(\tilde{\tilde{X}}) = \begin{pmatrix} \sigma_{11}I_n & \sigma_{12}I_n & \cdots & \sigma_{1p}I_n \\ \sigma_{21}I_n & \sigma_{22}I_n & \cdots & \sigma_{2p}I_n \\ \vdots & & \ddots & \vdots \\ \sigma_{p1}I_n & \sigma_{p2}I_n & \cdots & \sigma_{pp}I_n \end{pmatrix}$$

where each block is $n \times n$, $\Sigma = \{\sigma_{ij}\}$.

15. The unit sphere in R^n is the set $\{x | x \in R^n, \|x\| = 1\} = \mathcal{X}$. A random vector X with values in $\mathcal{X}$ has a *uniform* distribution on $\mathcal{X}$ if $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ for all $\Gamma \in \mathcal{O}_n$. (There is one and only one uniform distribution on $\mathcal{X}$—this is discussed in detail in Chapters 6 and 7.)

(i) Show that $\mathcal{E}X = 0$ and $\operatorname{Cov}(X) = (1/n)I_n$.

(ii) Let X_1 be the first coordinate of X and let $\dot{X} \in R^{n-1}$ be the remaining $n - 1$ coordinates. What is the best affine predictor of X_1 based on $\dot{X}$? How would you predict X_1 on the basis of $\dot{X}$?

16. Show that the linear transformation T_2 in Proposition 2.24 is $\Sigma \square \Sigma$ where $\square$ denotes the outer product of the vector space $(M_s, \langle \cdot, \cdot \rangle)$. Here, $\langle \cdot, \cdot \rangle$ is the natural inner product on $\mathcal{L}(V, V)$.

17. Suppose $X \in R^2$ has coordinates X_1 and X_2 that are independent with a standard normal distribution. Let $S = XX'$ and denote the elements of S by s_{11}, s_{22}, and $s_{12} = s_{21}$.

(i) What is the covariance matrix of

$$\begin{pmatrix} s_{11} \\ s_{12} \\ s_{22} \end{pmatrix} \in R^3?$$

(ii) Regard S as a random vector in $(\mathcal{S}_2, \langle \cdot, \cdot \rangle)$ (see Problem 12). What is $\operatorname{Cov}(S)$ in the space $(\mathcal{S}_2, \langle \cdot, \cdot \rangle)$?

(iii) How do you reconcile your answers to (i) and (ii)?

NOTES AND REFERENCES

1. In the first two sections of this chapter, we have simply translated well known coordinate space results into their inner product space versions. The coordinate space results can be found in Billingsley (1979). The inner product space versions were used by Kruskal (1961) in his work on missing and extra values in analysis of variance problems.

2. In the third section, topics with multivariate flavor emerge. The reader may find it helpful to formulate coordinate versions of each proposition. If nothing else, this exercise will soon explain my acquired preference for vector space, as opposed to coordinate, methods and notation.

3. Proposition 2.14 is a special case of Schur's Lemma—a basic result in group representation theory. The book by Serre (1977) is an excellent place to begin a study of group representations.