Appendix B
Matrix Results

This appendix reviews standard ideas in matrix theory with emphasis given to important results that are less commonly taught in a junior/senior level linear algebra course. The appendix begins with basic definitions and results. A section devoted to eigenvalues and their applications follows. This section contains a number of standard definitions, but it also contains a number of very specific results that are unlikely to be familiar to people with only an undergraduate background in linear algebra. The third section is devoted to an intense (brief but detailed) examination of projections and their properties. The appendix closes with some miscellaneous results, some results on Kronecker products and Vec operators, and an introduction to tensors.

B.1 Basic Ideas

Definition B.1. Any matrix with the same number of rows and columns is called a *square matrix*.

Definition B.2. Let $A = [a_{ij}]$ be a matrix. The *transpose* of A, written A', is the matrix $A' = [b_{ij}]$, where $b_{ij} = a_{ji}$.

Definition B.3. If $A = A'$, then A is called *symmetric*. Note that only square matrices can be symmetric.

Definition B.4. If $A = [a_{ij}]$ is a square matrix and $a_{ij} = 0$ for $i \neq j$, then A is a *diagonal matrix*. If $\lambda_1, \ldots, \lambda_n$ are scalars, then $D(\lambda_j)$ and $\mathrm{Diag}(\lambda_j)$ are used to indicate an $n \times n$ matrix $D = [d_{ij}]$ with $d_{ij} = 0$, $i \neq j$, and $d_{ii} = \lambda_i$. If $\lambda \equiv (\lambda_1, \ldots, \lambda_n)'$, then

$D(\lambda) \equiv D(\lambda_j)$. A diagonal matrix with all 1s on the diagonal is called an *identity matrix* and is denoted I. Occasionally, I_n is used to denote an $n \times n$ identity matrix.

If $A = [a_{ij}]$ is $n \times p$ and $B = [b_{ij}]$ is $n \times q$, we can write an $n \times (p+q)$ matrix $C = [A, B]$, where $c_{ij} = a_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, and $c_{ij} = b_{i,j-p}$, $i = 1, \ldots, n$, $j = p+1, \ldots, p+q$. This notation can be extended in obvious ways, e.g., $C' = \begin{bmatrix} A' \\ B' \end{bmatrix}$.

Definition B.5. Let $A = [a_{ij}]$ be an $r \times c$ matrix and $B = [b_{ij}]$ be an $s \times d$ matrix. The *Kronecker product* of A and B, written $A \otimes B$, is an $r \times c$ matrix of $s \times d$ matrices. The matrix in the ith row and jth column is $a_{ij}B$. In total, $A \otimes B$ is an $rs \times cd$ matrix.

Definition B.6. Let A be an $r \times c$ matrix. Write $A = [A_1, A_2, \ldots, A_c]$, where A_i is the ith column of A. Then the *Vec* operator stacks the columns of A into an $rc \times 1$ vector; thus,

$$[\mathrm{Vec}(A)]' = [A_1', A_2', \ldots, A_c'].$$

EXAMPLE B.7.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix},$$

$$A \otimes B = \begin{bmatrix} 1\begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix} & 4\begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix} \\ 2\begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix} & 5\begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 4 & 12 \\ 0 & 4 & 0 & 16 \\ 2 & 6 & 5 & 15 \\ 0 & 8 & 0 & 20 \end{bmatrix},$$

$$\mathrm{Vec}(A) = [1, 2, 4, 5]'.$$

Definition B.8. Let A be an $n \times n$ matrix. A is *nonsingular* if there exists a matrix A^{-1} such that $A^{-1}A = I = AA^{-1}$. If no such matrix exists, then A is singular. If A^{-1} exists, it is called the *inverse* of A.

Theorem B.9. An $n \times n$ matrix A is nonsingular if and only if $r(A) = n$, i.e., the columns of A form a basis for $\mathbf{R}^n$.

Corollary B.10. $A_{n \times n}$ is singular if and only if there exists $x \neq 0$ such that $Ax = 0$.

For any matrix A, the set of all x such that $Ax = 0$ is easily seen to be a vector space.

Definition B.11. The set of all x such that $Ax = 0$ is called the *null space* of A.

Theorem B.12. If A is $n \times n$ and $r(A) = r$, then the null space of A has rank $n - r$.

B.2 Eigenvalues and Related Results

The material in this section deals with eigenvalues and eigenvectors either in the statements of the results or in their proofs. Again, this is meant to be a brief review of important concepts; but, in addition, there are a number of specific results that may be unfamiliar.

Definition B.13. The scalar λ is an *eigenvalue* of $A_{n\times n}$ if $A - \lambda I$ is singular. λ is an eigenvalue of *multiplicity* s if the rank of the null space of $A - \lambda I$ is s. A nonzero vector x is an *eigenvector* of A corresponding to the eigenvalue λ if x is in the null space of $A - \lambda I$, i.e., if $Ax = \lambda x$. Eigenvalues are also called *singular values* and *characteristic roots*.

For example,

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Combining the two equations gives

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that if $\lambda \neq 0$ is an eigenvalue of A, the eigenvectors corresponding to λ (along with the vector 0) form a subspace of $C(A)$. For example, if $Ax_1 = \lambda x_1$ and $Ax_2 = \lambda x_2$, then $A(x_1 + x_2) = \lambda(x_1 + x_2)$, so the set of eigenvectors is closed under vector addition. Similarly, it is closed under scalar multiplication, so it forms a subspace (except that eigenvectors cannot be 0 and every subspace contains 0). If $\lambda = 0$, the subspace is the null space of A.

If A is a symmetric matrix, and γ and λ are distinct eigenvalues, then the eigenvectors corresponding to λ and γ are orthogonal. To see this, let x be an eigenvector for λ and y an eigenvector for γ. Then $\lambda x'y = x'Ay = \gamma x'y$, which can happen only if $\lambda = \gamma$ or if $x'y = 0$. Since λ and γ are distinct, we have $x'y = 0$.

Let $\lambda_1, \ldots, \lambda_r$ be the distinct nonzero eigenvalues of a symmetric matrix A with respective multiplicities $s(1), \ldots, s(r)$. Let $v_{i1}, \ldots, v_{is(i)}$ be a basis for the space of eigenvectors of λ_i. We want to show that $v_{11}, v_{12}, \ldots, v_{rs(r)}$ is a basis for $C(A)$. Suppose $v_{11}, v_{12}, \ldots, v_{rs(r)}$ is not a basis. Since $v_{ij} \in C(A)$ and the v_{ij}s are linearly independent, we can pick $x \in C(A)$ with $x \perp v_{ij}$ for all i and j. Note that since $Av_{ij} = \lambda_i v_{ij}$, we have $(A)^p v_{ij} = (\lambda_i)^p v_{ij}$. In particular, $x'(A)^p v_{ij} = x'(\lambda_i)^p v_{ij} = (\lambda_i)^p x' v_{ij} = 0$, so $A^p x \perp v_{ij}$ for any i, j, and p. The vectors x, Ax, $A^2x, \ldots$ cannot all be linearly independent, so there exists a smallest value $k \leq n$ such that

$$A^k x + b_{k-1}A^{k-1}x + \cdots + b_0 x = 0.$$

Since there is a solution to this, for some real number μ we can write the equation as

$$(A - \mu I)\left(A^{k-1}x + \gamma_{k-2}A^{k-2}x + \cdots + \gamma_0 x\right) = 0,$$

and μ is an eigenvalue. (See Exercise B.1.) An eigenvector for μ is $y = A^{k-1}x + \gamma_{k-2}A^{k-2}x + \cdots + \gamma_0 x$. Clearly, $y \perp v_{ij}$ for any i and j. Since k was chosen as the smallest value to get linear dependence, we have $y \neq 0$. If $\mu \neq 0$, y is an eigenvector that does not correspond to any of $\lambda_1, \ldots, \lambda_r$, a contradiction. If $\mu = 0$, we have $Ay = 0$; and since A is symmetric, y is a vector in $C(A)$ that is orthogonal to every other vector in $C(A)$, i.e., $y'y = 0$ but $y \neq 0$, a contradiction. We have proven

Theorem B.14. If A is a symmetric matrix, then there exists a basis for $C(A)$ consisting of eigenvectors of nonzero eigenvalues. If λ is a nonzero eigenvalue of multiplicity s, then the basis will contain s eigenvectors for λ.

If λ is an eigenvalue of A with multiplicity s, then we can think of λ as being an eigenvalue s times. With this convention, the rank of A is the number of nonzero eigenvalues. The total number of eigenvalues is n if A is an $n \times n$ matrix.

For a symmetric matrix A, if we use eigenvectors corresponding to the zero eigenvalue, we can get a basis for $\mathbf{R}^n$ consisting of eigenvectors. We already have a basis for $C(A)$, and the eigenvectors of 0 are the null space of A. For A symmetric, $C(A)$ and the null space of A are orthogonal complements. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of a symmetric matrix A. Let $v_1, \ldots, v_n$ denote a basis of eigenvectors for $\mathbf{R}^n$, with v_i being an eigenvector for λ_i for any i.

Theorem B.15. If A is symmetric, there exists an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors of A.

PROOF. Assume $\lambda_{i1} = \cdots = \lambda_{ik}$ are all the λ_is equal to any particular value λ, and let $v_{i1}, \ldots, v_{ik}$ be a basis for the space of eigenvectors for λ. By Gram–Schmidt there exists an orthonormal basis $w_{i1}, \ldots, w_{ik}$ for the space of eigenvectors corresponding to λ. If we do this for each distinct eigenvalue, we get a collection of orthonormal sets that form a basis for $\mathbf{R}^n$. Since, as we have seen, for $\lambda_i \neq \lambda_j$, any eigenvector for λ_i is orthogonal to any eigenvector for λ_j, the basis is orthonormal. □

Definition B.16. A square matrix P is *orthogonal* if $P' = P^{-1}$. Note that if P is orthogonal, so is P'.

Some examples of orthogonal matrices are

$$P_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \\ \sqrt{2} & \sqrt{3} & 1 \end{bmatrix}, \quad P_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

$$P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Theorem B.17. $P_{n\times n}$ is orthogonal if and only if the columns of P form an orthonormal basis for $\mathbf{R}^n$.

PROOF. $\Leftarrow$ It is clear that if the columns of P form an orthonormal basis for $\mathbf{R}^n$, then $P'P = I$.

$\Rightarrow$ Since P is nonsingular, the columns of P form a basis for $\mathbf{R}^n$. Since $P'P = I$, the basis is orthonormal. □

Corollary B.18. $P_{n\times n}$ is orthogonal if and only if the rows of P form an orthonormal basis for $\mathbf{R}^n$.

PROOF. P is orthogonal if and only if P' is orthogonal if and only if the columns of P' are an orthonormal basis if and only if the rows of P are an orthonormal basis. □

Theorem B.19. If A is an $n \times n$ symmetric matrix, then there exists an orthogonal matrix P such that $P'AP = \text{Diag}(\lambda_i)$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

PROOF. Let $v_1, v_2, \ldots, v_n$ be an orthonormal set of eigenvectors of A corresponding, respectively, to $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $P = [v_1, \ldots, v_n]$. Then

$$\begin{aligned} P'AP &= \begin{bmatrix} v_1' \\ \vdots \\ v_n' \end{bmatrix} [Av_1, \ldots, Av_n] \\ &= \begin{bmatrix} v_1' \\ \vdots \\ v_n' \end{bmatrix} [\lambda_1 v_1, \ldots, \lambda_n v_n] \\ &= \begin{bmatrix} \lambda_1 v_1' v_1 & \ldots & \lambda_n v_1' v_n \\ \vdots & \ddots & \vdots \\ \lambda_1 v_n' v_1 & \ldots & \lambda_n v_n' v_n \end{bmatrix} \\ &= \text{Diag}(\lambda_i). \end{aligned}$$

□

The *singular value decomposition* for a symmetric matrix is given by the following corollary.

Corollary B.20. $A = PD(\lambda_i)P'$.

For example, using results illustrated earlier,

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

Definition B.21. A symmetric matrix A is *positive (nonnegative) definite* if, for any nonzero vector $v \in \mathbf{R}^n$, $v'Av$ is positive (nonnegative).

Theorem B.22. A is nonnegative definite if and only if there exists a square matrix Q such that $A = QQ'$.

PROOF. $\Rightarrow$ We know that there exists P orthogonal with $P'AP = \text{Diag}(\lambda_i)$. The λ_is must all be nonnegative, because if $e'_j = (0,\dots,0,1,0,\dots,0)$ with the 1 in the jth place and we let $v = Pe_j$, then $0 \le v'Av = e'_j\text{Diag}(\lambda_i)e_j = \lambda_j$. Let $Q = P\text{Diag}\left(\sqrt{\lambda_i}\right)$. Then, since $P\text{Diag}(\lambda_i)P' = A$, we have

$$QQ' = P\text{Diag}(\lambda_i)P' = A.$$

$\Leftarrow$ If $A = QQ'$, then $v'Av = (Q'v)'(Q'v) \ge 0$. □

Corollary B.23. A is positive definite if and only if Q is nonsingular for any choice of Q.

PROOF. There exists $v \ne 0$ such that $v'Av = 0$ if and only if there exists $v \ne 0$ such that $Q'v = 0$, which occurs if and only if Q' is singular. The contrapositive of this is that $v'Av > 0$ for all $v \ne 0$ if and only if Q' is nonsingular. □

Theorem B.24. If A is an $n \times n$ nonnegative definite matrix with nonzero eigenvalues $\lambda_1,\dots,\lambda_r$, then there exists an $n \times r$ matrix $Q = Q_1Q_2^{-1}$ such that Q_1 has orthonormal columns, $C(Q_1) = C(A)$, Q_2 is diagonal and nonsingular, and $Q'AQ = I$.

PROOF. Let $v_1,\dots,v_n$ be an orthonormal basis of eigenvectors with $v_1,\dots,v_r$ corresponding to $\lambda_1,\dots,\lambda_r$. Let $Q_1 = [v_1,\dots,v_r]$. By an argument similar to that in the proof of Theorem B.19, $Q'_1AQ_1 = \text{Diag}(\lambda_i)$, $i = 1,\dots,r$. Now take $Q_2 = \text{Diag}(\sqrt{\lambda_i})$ and $Q = Q_1Q_2^{-1}$. □

Corollary B.25. Let $W = Q_1Q_2$. Then $WW' = A$.

PROOF. Since $Q'AQ = Q_2^{-1}Q_1'AQ_1Q_2^{-1} = I$ and Q_2 is symmetric, $Q_1'AQ_1 = Q_2Q_2'$. Multiplying gives

$$Q_1Q_1'AQ_1Q_1' = (Q_1Q_2)(Q_2'Q_1') = WW'.$$

But Q_1Q_1' is a perpendicular projection matrix onto $C(A)$, so $Q_1Q_1'AQ_1Q_1' = A$ (cf. Definition B.31 and Theorem B.35). $\square$

Corollary B.26. $AQQ'A = A$ and $QQ'AQQ' = QQ'$.

PROOF. $AQQ'A = WW'QQ'WW' = WQ_2Q_1'Q_1Q_2^{-1}Q_2^{-1}Q_1'Q_1Q_2W' = A$. Moreover, $QQ'AQQ' = QQ'WW'QQ' = QQ_2^{-1}Q_1'Q_1Q_2Q_2Q_1'Q_1Q_2^{-1}Q' = QQ'$. $\square$

Definition B.27. Let $A = [a_{ij}]$ be an $n \times n$ matrix. The *trace* of A is $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$.

Theorem B.28. For matrices $A_{r\times s}$ and $B_{s\times r}$, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

PROOF. See Exercise B.8. $\square$

Theorem B.29. If A is a symmetric matrix, $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

PROOF. $A = PD(\lambda_i)P'$ with P orthogonal

$$\begin{aligned}\mathrm{tr}(A) &= \mathrm{tr}(PD(\lambda_i)P') = \mathrm{tr}(D(\lambda_i)P'P)\\ &= \mathrm{tr}(D(\lambda_i)) = \sum_{i=1}^n \lambda_i.\end{aligned}$$ $\square$

To illustrate, we saw earlier that the matrix $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ had eigenvalues of 3 and 1. In fact, a stronger result than Theorem B.29 is true. We give it without proof.

Theorem B.30. $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Moreover, the determinant of A is $\det(V) = \prod_{i=1}^n \lambda_i$.

B.3 Projections

This section is devoted primarily to a discussion of perpendicular projection operators. It begins with their definition, some basic properties, and two important characterizations: Theorems B.33 and B.35. A third important characterization, Theorem B.44, involves generalized inverses. Generalized inverses are defined, briefly studied, and applied to projection operators. The section continues with the examination of the relationships between two perpendicular projection operators and closes with discussions of the Gram–Schmidt theorem, eigenvalues of projection operators, and oblique (nonperpendicular) projection operators.

We begin by defining a *perpendicular projection operator (ppo)* onto an arbitrary space. To be consistent with later usage, we denote the arbitrary space $C(X)$ for some matrix X.

Definition B.31. M is a perpendicular projection operator (matrix) onto $C(X)$ if and only if

(i) $v \in C(X)$ implies $Mv = v$ (projection),
(ii) $w \perp C(X)$ implies $Mw = 0$ (perpendicularity).

For example, consider the subspace of $\mathbf{R}^2$ determined by vectors of the form $(2a, a)'$. It is not difficult to see that the orthogonal complement of this subspace consists of vectors of the form $(b, -2b)'$. The perpendicular projection operator onto the $(2a, a)'$ subspace is

$$M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}.$$

To verify this note that

$$M\begin{pmatrix} 2a \\ a \end{pmatrix} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}\begin{pmatrix} 2a \\ a \end{pmatrix} = \begin{pmatrix} (0.8)2a + 0.4a \\ (0.4)2a + 0.2a \end{pmatrix} = \begin{pmatrix} 2a \\ a \end{pmatrix}$$

and

$$M\begin{pmatrix} b \\ -2b \end{pmatrix} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}\begin{pmatrix} b \\ -2b \end{pmatrix} = \begin{pmatrix} 0.8b + 0.4(-2b) \\ 0.4b + 0.2(-2b) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Notationally, M is used to indicate the ppo onto $C(X)$. If A is another matrix, M_A denotes the ppo onto $C(A)$. Thus, $M \equiv M_X$. When using X with a subscript, say 0, write the ppo onto $C(X_0)$ as $M_0 \equiv M_{X_0}$.

Proposition B.32. If M is a perpendicular projection operator onto $C(X)$, then $C(M) = C(X)$.

PROOF. See Exercise B.2. □

Note that both columns of

$$M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$$

have the form $(2a, a)'$.

Theorem B.33. M is a perpendicular projection operator on $C(M)$ if and only if $MM = M$ and $M' = M$.

PROOF. $\Rightarrow$ Write $v = v_1 + v_2$, where $v_1 \in C(M)$ and $v_2 \perp C(M)$, and let $w = w_1 + w_2$ with $w_1 \in C(M)$ and $w_2 \perp C(M)$. Since $(I-M)v = (I-M)v_2 = v_2$ and $Mw = Mw_1 = w_1$, we get

$$w'M'(I-M)v = w_1'M'(I-M)v_2 = w_1'v_2 = 0.$$

This is true for any v and w, so we have $M'(I-M) = 0$ or $M' = M'M$. Since $M'M$ is symmetric, M' must also be symmetric, and this implies that $M = MM$.

$\Leftarrow$ If $M^2 = M$ and $v \in C(M)$, then since $v = Mb$ we have $Mv = MMb = Mb = v$. If $M' = M$ and $w \perp C(M)$, then $Mw = M'w = 0$ because the columns of M are in $C(M)$. □

In our example,

$$MM = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = M$$

and

$$M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = M'.$$

Proposition B.34. Perpendicular projection operators are unique.

PROOF. Let M and P be perpendicular projection operators onto some space $\mathscr{M}$. Let $v \in \mathbf{R}^n$ and write $v = v_1 + v_2$, $v_1 \in \mathscr{M}$, $v_2 \perp \mathscr{M}$. Since v is arbitrary and $Mv = v_1 = Pv$, we have $M = P$. □

For any matrix X, we will now find two ways to characterize the perpendicular projection operator onto $C(X)$. The first method depends on the Gram–Schmidt theorem; the second depends on the concept of a generalized inverse.

Theorem B.35. Let $o_1, \ldots, o_r$ be an orthonormal basis for $C(X)$, and let $O = [o_1, \ldots, o_r]$. Then $OO' = \sum_{i=1}^r o_i o_i'$ is the perpendicular projection operator onto $C(X)$.

PROOF. OO' is symmetric and $OO'OO' = OI_rO' = OO'$; so, by Theorem B.33, it only remains to show that $C(OO') = C(X)$. Clearly $C(OO') \subset C(O) = C(X)$. On the other hand, if $v \in C(O)$, then $v = Ob$ for some vector $b \in \mathbf{R}^r$ and $v = Ob = OI_rb = OO'Ob$; so clearly $v \in C(OO')$. □

For example, to find the perpendicular projection operator for vectors of the form $(2a,a)'$, we can find an orthonormal basis. The space has rank 1 and to normalize $(2a,a)'$, we must have

$$1 = (2a,a)' \begin{pmatrix} 2a \\ a \end{pmatrix} = 4a^2 + a^2 = 5a^2;$$

so $a^2 = 1/5$ and $a = \pm 1/\sqrt{5}$. If we take $(2/\sqrt{5}, 1/\sqrt{5})'$ as our orthonormal basis, then

$$M = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix} (2/\sqrt{5}, 1/\sqrt{5}) = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix},$$

as was demonstrated earlier.

One use of Theorem B.35 is that, given a matrix X, one can use the Gram–Schmidt theorem to get an orthonormal basis for $C(X)$ and thus obtain the perpendicular projection operator.

We now examine properties of generalized inverses. Generalized inverses are a generalization on the concept of the inverse of a matrix. Although the most common use of generalized inverses is in solving systems of linear equations, our interest lies primarily in their relationship to projection operators. The discussion below is given for an arbitrary matrix A.

Definition B.36. A *generalized inverse* of a matrix A is any matrix G such that $AGA = A$. The notation A^- is used to indicate a generalized inverse of A.

Theorem B.37. If A is nonsingular, the unique generalized inverse of A is A^{-1}.

PROOF. $AA^{-1}A = IA = A$, so A^{-1} is a generalized inverse. If $AA^-A = A$, then $AA^- = AA^-AA^{-1} = AA^{-1} = I$; so A^- is the inverse of A. □

Theorem B.38. For any symmetric matrix A, there exists a generalized inverse of A.

PROOF. There exists P orthogonal so that $P'AP = D(\lambda_i)$ and $A = PD(\lambda_i)P'$. Let

$$\gamma_i = \begin{cases} 1/\lambda_i, & \text{if } \lambda_i \neq 0 \\ 0, & \text{if } \lambda_i = 0 \end{cases},$$

and $G = PD(\gamma_i)P'$. We now show that G is a generalized inverse of A. P is orthogonal, so $P'P = I$ and

$$\begin{aligned} AGA &= PD(\lambda_i)P'PD(\gamma_i)P'PD(\lambda_i)P' \\ &= PD(\lambda_i)D(\gamma_i)D(\lambda_i)P' \\ &= PD(\lambda_i)P' \\ &= A. \end{aligned}$$

□

Although this is the only existence result we really need, later we will show that generalized inverses exist for arbitrary matrices.

Theorem B.39. If G_1 and G_2 are generalized inverses of A, then so is G_1AG_2.

PROOF. $A(G_1AG_2)A = (AG_1A)G_2A = AG_2A = A$. □

For A symmetric, A^- need not be symmetric.

EXAMPLE B.40. Consider the matrix

$$\begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix}.$$

It has a generalized inverse

$$\begin{bmatrix} 1/a & -1 \\ 1 & 0 \end{bmatrix},$$

and in fact, by considering the equation

$$\begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix} \begin{bmatrix} r & s \\ t & u \end{bmatrix} \begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix} = \begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix},$$

it can be shown that if $r = 1/a$, then any solution of $at + as + bu = 0$ gives a generalized inverse.

Corollary B.41. For a symmetric matrix A, there exists A^- such that $A^-AA^- = A^-$ and $(A^-)' = A^-$.

PROOF. Take A^- as the generalized inverse in the proof of Theorem B.38. Clearly, $A^- = PD(\gamma_i)P'$ is symmetric and

$$A^-AA^- = PD(\gamma_i)P'PD(\lambda_i)P'PD(\gamma_i)P' = PD(\gamma_i)D(\lambda_i)D(\gamma_i)P' = PD(\gamma_i)P' = A^-.$$

□

Definition B.42. A generalized inverse A^- for a matrix A that has the property $A^-AA^- = A^-$ is said to be *reflexive*.

Corollary B.41 establishes the existence of a reflexive generalized inverse for any symmetric matrix. Note that Corollary B.26 previously established the existence of a reflexive generalized inverse for any nonnegative definite matrix.

Generalized inverses are of interest in that they provide an alternative to the characterization of perpendicular projection matrices given in Theorem B.35. The two results immediately below characterize the perpendicular projection matrix onto $C(X)$.

Lemma B.43. If G and H are generalized inverses of $(X'X)$, then

(i) $XGX'X = XHX'X = X$,
(ii) $XGX' = XHX'$.

PROOF. For $v \in \mathbf{R}^n$, let $v = v_1 + v_2$ with $v_1 \in C(X)$ and $v_2 \perp C(X)$. Also let $v_1 = Xb$ for some vector b. Then

$$v'XGX'X = v_1'XGX'X = b'(X'X)G(X'X) = b'(X'X) = v'X.$$

Since v and G are arbitrary, we have shown (i).

To see (ii), observe that for the arbitrary vector v above,

$$XGX'v = XGX'Xb = XHX'Xb = XHX'v. \qquad \square$$

Since $X'X$ is symmetric, there exists a generalized inverse $(X'X)^-$ that is symmetric. For this generalized inverse, $X(X'X)^-X'$ is symmetric; so, by the above lemma, $X(X'X)^-X'$ must be symmetric for any choice of $(X'X)^-$.

Theorem B.44. $X(X'X)^-X'$ is the perpendicular projection operator onto $C(X)$.

PROOF. We need to establish conditions (i) and (ii) of Definition B.31. (i) For $v \in C(X)$, write $v = Xb$, so by Lemma B.43, $X(X'X)^-X'v = X(X'X)^-X'Xb = Xb = v$. (ii) If $w \perp C(X)$, $X(X'X)^-X'w = 0$. $\square$

For example, one spanning set for the subspace of vectors with the form $(2a, a)'$ is $(2,1)'$. It follows that

$$M = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \left[(2,1) \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right]^{-1} (2,1) = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix},$$

as was shown earlier.

The next five results examine the relationships between two perpendicular projection matrices.

Theorem B.45. Let M_1 and M_2 be perpendicular projection matrices on $\mathbf{R}^n$. $(M_1 + M_2)$ is the perpendicular projection matrix onto $C(M_1, M_2)$ if and only if $C(M_1) \perp C(M_2)$.

PROOF. $\Leftarrow$ If $C(M_1) \perp C(M_2)$, then $M_1M_2 = M_2M_1 = 0$. Because

$$(M_1 + M_2)^2 = M_1^2 + M_2^2 + M_1M_2 + M_2M_1 = M_1^2 + M_2^2 = M_1 + M_2$$

and

$$(M_1+M_2)' = M_1' + M_2' = M_1 + M_2,$$

M_1+M_2 is the perpendicular projection matrix onto $C(M_1+M_2)$. Clearly $C(M_1+M_2) \subset C(M_1,M_2)$. To see that $C(M_1,M_2) \subset C(M_1+M_2)$, write $v = M_1b_1 + M_2b_2$. Then, because $M_1M_2 = M_2M_1 = 0$, $(M_1+M_2)v = v$. Thus, $C(M_1,M_2) = C(M_1+M_2)$.

$\Rightarrow$ If M_1+M_2 is a perpendicular projection matrix, then

$$\begin{aligned}(M_1+M_2) &= (M_1+M_2)^2 = M_1^2 + M_2^2 + M_1M_2 + M_2M_1 \\ &= M_1 + M_2 + M_1M_2 + M_2M_1.\end{aligned}$$

Thus, $M_1M_2 + M_2M_1 = 0$.

Multiplying by M_1 gives $0 = M_1^2M_2 + M_1M_2M_1 = M_1M_2 + M_1M_2M_1$ and thus $-M_1M_2M_1 = M_1M_2$. Since $-M_1M_2M_1$ is symmetric, so is M_1M_2. This gives $M_1M_2 = (M_1M_2)' = M_2M_1$, so the condition $M_1M_2 + M_2M_1 = 0$ becomes $2(M_1M_2) = 0$ or $M_1M_2 = 0$. By symmetry, this says that the columns of M_1 are orthogonal to the columns of M_2. □

Theorem B.46. If M_1 and M_2 are symmetric, $C(M_1) \perp C(M_2)$, and (M_1+M_2) is a perpendicular projection matrix, then M_1 and M_2 are perpendicular projection matrices.

PROOF.

$$(M_1+M_2) = (M_1+M_2)^2 = M_1^2 + M_2^2 + M_1M_2 + M_2M_1.$$

Since M_1 and M_2 are symmetric with $C(M_1) \perp C(M_2)$, we have $M_1M_2 + M_2M_1 = 0$ and $M_1 + M_2 = M_1^2 + M_2^2$. Rearranging gives $M_2 - M_2^2 = M_1^2 - M_1$, so $C(M_2 - M_2^2) = C(M_1^2 - M_1)$. Now $C(M_2 - M_2^2) \subset C(M_2)$ and $C(M_1^2 - M_1) \subset C(M_1)$, so $C(M_2 - M_2^2) \perp C(M_1^2 - M_1)$. The only way a vector space can be orthogonal to itself is if it consists only of the zero vector. Thus, $M_2 - M_2^2 = M_1^2 - M_1 = 0$, and $M_2 = M_2^2$ and $M_1 = M_1^2$. □

Theorem B.47. Let M and M_0 be perpendicular projection matrices with $C(M_0) \subset C(M)$. Then $M - M_0$ is a perpendicular projection matrix.

PROOF. Since $C(M_0) \subset C(M)$, $MM_0 = M_0$ and, by symmetry, $M_0M = M_0$. Checking the conditions of Theorem B.33, we see that $(M-M_0)^2 = M^2 - MM_0 - M_0M + M_0^2 = M - M_0 - M_0 + M_0 = M - M_0$, and $(M-M_0)' = M - M_0$. □

Theorem B.48. Let M and M_0 be perpendicular projection matrices with $C(M_0) \subset C(M)$. Then $C(M-M_0)$ is the orthogonal complement of $C(M_0)$ with respect to $C(M)$, i.e., $C(M-M_0) = C(M_0)^{\perp}_{C(M)}$.

PROOF. $C(M-M_0) \perp C(M_0)$ because $(M-M_0)M_0 = MM_0 - M_0^2 = M_0 - M_0 = 0$. Thus, $C(M-M_0)$ is contained in the orthogonal complement of $C(M_0)$ with respect to $C(M)$. If $x \in C(M)$ and $x \perp C(M_0)$, then $x = Mx = (M-M_0)x + M_0x = (M-M_0)x$. Thus, $x \in C(M-M_0)$, so the orthogonal complement of $C(M_0)$ with respect to $C(M)$ is contained in $C(M-M_0)$. □

Corollary B.49. $r(M) = r(M_0) + r(M-M_0)$.

One particular application of these results involves I, the perpendicular projection operator onto $\mathbf{R}^n$. For any other perpendicular projection operator M, $I-M$ is the perpendicular projection operator onto the orthogonal complement of $C(M)$ with respect to $\mathbf{R}^n$.

For example, the subspace of vectors with the form $(2a,a)'$ has an orthogonal complement consisting of vectors with the form $(b,-2b)'$. With M as given earlier,

$$I-M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix}.$$

Note that

$$(I-M)\begin{pmatrix} b \\ -2b \end{pmatrix} = \begin{pmatrix} b \\ -2b \end{pmatrix} \quad \text{and} \quad (I-M)\begin{pmatrix} 2a \\ a \end{pmatrix} = 0;$$

so by definition $I-M$ is the perpendicular projection operator onto the space of vectors with the form $(b,-2b)'$.

At this point, we examine the relationship between perpendicular projection operations and the Gram–Schmidt theorem (Theorem A.12). Recall that in the Gram–Schmidt theorem, $x_1,\ldots,x_r$ denotes the original basis and $y_1,\ldots,y_r$ denotes the orthonormal basis. Let

$$M_s = \sum_{i=1}^{s} y_i y_i'.$$

Applying Theorem B.35, M_s is the ppo onto $C(x_1,\ldots,x_s)$. Now define

$$w_{s+1} = (I-M_s)x_{s+1}.$$

Thus, w_{s+1} is the projection of x_{s+1} onto the orthogonal complement of $C(x_1,\ldots,x_s)$. Finally, y_{s+1} is just w_{s+1} normalized.

Consider the eigenvalues of a perpendicular projection operator M. Let $v_1,\ldots,v_r$ be a basis for $C(M)$. Then $Mv_i = v_i$, so v_i is an eigenvector of M with eigenvalue 1. In fact, 1 is an eigenvalue of M with multiplicity r. Now, let $w_1,\ldots,w_{n-r}$ be a basis for $C(M)^\perp$. $Mw_j = 0$, so 0 is an eigenvalue of M with multiplicity $n-r$. We have completely characterized the n eigenvalues of M. Since $\text{tr}(M)$ equals the sum of the eigenvalues, we have $\text{tr}(M) = r(M)$.

In fact, if A is an $n \times n$ matrix with $A^2 = A$, any basis for $C(A)$ is a basis for the space of eigenvectors for the eigenvalue 1. The null space of A is the space of eigenvectors for the eigenvalue 0. The rank of A and the rank of the null space of A

add to n, and A has n eigenvalues, so all the eigenvalues are accounted for. Again, $\mathrm{tr}(A) = r(A)$.

Definition B.50.

(a) If A is a square matrix with $A^2 = A$, then A is called *idempotent*.
(b) Let $\mathscr{N}$ and $\mathscr{M}$ be two spaces with $\mathscr{N} \cap \mathscr{M} = \{0\}$ and $r(\mathscr{N}) + r(\mathscr{M}) = n$. The square matrix A is a *projection operator* onto $\mathscr{N}$ along $\mathscr{M}$ if 1) $Av = v$ for any $v \in \mathscr{N}$, and 2) $Aw = 0$ for any $w \in \mathscr{M}$.

If the square matrix A has the property that $Av = v$ for any $v \in C(A)$, then A is the projection operator (matrix) onto $C(A)$ along $C(A')^{\perp}$. (Note that $C(A')^{\perp}$ is the null space of A.) It follows immediately that if A is idempotent, then A is a projection operator onto $C(A)$ along $\mathscr{N}(A) = C(A')^{\perp}$.

The uniqueness of projection operators can be established like it was for perpendicular projection operators. Note that $x \in \mathbf{R}^n$ can be written uniquely as $x = v + w$ for $v \in \mathscr{N}$ and $w \in \mathscr{M}$. To see this, take basis matrices for the two spaces, say N and M, respectively. The result follows from observing that $[N, M]$ is a basis matrix for $\mathbf{R}^n$. Because of the rank conditions, $[N, M]$ is an $n \times n$ matrix. It is enough to show that the columns of $[N, M]$ must be linearly independent.

$$0 = [N, M] \begin{bmatrix} b \\ c \end{bmatrix} = Nb + Mc$$

implies $Nb = M(-c)$ which, since $\mathscr{N} \cap \mathscr{M} = \{0\}$, can only happen when $Nb = 0 = M(-c)$, which, because they are basis matrices, can only happen when $b = 0 = (-c)$, which implies that $\begin{bmatrix} b \\ c \end{bmatrix} = 0$, and we are done.

Any projection operator that is not a perpendicular projection is referred to as an *oblique projection operator*.

To show that a matrix A is a projection operator onto an arbitrary space, say $C(X)$, it is necessary to show that $C(A) = C(X)$ and that for $x \in C(X)$, $Ax = x$. A typical proof runs in the following pattern. First, show that $Ax = x$ for any $x \in C(X)$. This also establishes that $C(X) \subset C(A)$. To finish the proof, it suffices to show that $Av \in C(X)$ for any $v \in \mathbf{R}^n$ because this implies that $C(A) \subset C(X)$.

In this book, our use of the word "perpendicular" is based on the standard inner product, that defines Euclidean distance. In other words, for two vectors x and y, their inner product is $x'y$. By definition, the vectors x and y are orthogonal if their inner product is 0. In fact, for any two vectors x and y, let θ be the angle between x and y. Then $x'y = \sqrt{x'x}\sqrt{y'y}\cos\theta$. The length of a vector x is defined as the square root of the inner product of x with itself, i.e., $\|x\| \equiv \sqrt{x'x}$. The distance between two vectors x and y is the length of their difference, i.e., $\|x - y\|$.

These concepts can be generalized. For a positive definite matrix B, we can define an inner product between x and y as $x'By$. As before, x and y are orthogonal if their inner product is 0 and the length of x is the square root of its inner product with itself (now $\|x\|_B \equiv \sqrt{x'Bx}$). As argued above, any idempotent matrix is always a projection operator, but which one is the perpendicular projection operator depends on the inner product. As can be seen from Proposition 2.7.2 and Exercise 2.5, the matrix $X(X'BX)^{-}X'B$ is an oblique projection onto $C(X)$ for the standard inner product; but it is the perpendicular projection operator onto $C(X)$ with the inner product defined using the matrix B.

B.4 Miscellaneous Results

Proposition B.51. For any matrix X, $C(XX') = C(X)$.

PROOF. Clearly $C(XX') \subset C(X)$, so we need to show that $C(X) \subset C(XX')$. Let $x \in C(X)$. Then $x = Xb$ for some b. Write $b = b_0 + b_1$, where $b_0 \in C(X')$ and $b_1 \perp C(X')$. Clearly, $Xb_1 = 0$, so we have $x = Xb_0$. But $b_0 = X'd$ for some d; so $x = Xb_0 = XX'd$ and $x \in C(XX')$. □

Corollary B.52. For any matrix X, $r(XX') = r(X)$.

PROOF. See Exercise B.4. □

Corollary B.53. If $X_{n\times p}$ has $r(X) = p$, then the $p \times p$ matrix $X'X$ is nonsingular.

PROOF. See Exercise B.5. □

Proposition B.54. If B is nonsingular, $C(XB) = C(X)$.

PROOF. Clearly, $C(XB) \subset C(X)$. To see that $C(X) \subset C(XB)$, take $x \in C(X)$. It follows that for some vector b, $x = Xb$; so $x = XB(B^{-1}b) \in C(XB)$. □

It follows immediately from Proposition B.54 that the perpendicular projection operators onto $C(XB)$ and $C(X)$ are identical.

We now show that generalized inverses always exist.

Theorem B.55. For any matrix X, there exists a generalized inverse X^{-}.

PROOF. We know that $(X'X)^{-}$ exists. Set $X^{-} = (X'X)^{-}X'$. Then $XX^{-}X = X(X'X)^{-}X'X = X$ because $X(X'X)^{-}X'$ is a projection matrix onto $C(X)$. □

Note that for any X^-, the matrix XX^- is idempotent and hence a projection operator.

Proposition B.56. When all inverses exist,

$$[A+BCD]^{-1} = A^{-1} - A^{-1}B\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1}.$$

PROOF.

$$\begin{aligned}
&[A+BCD]\left[A^{-1} - A^{-1}B\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1}\right] \\
&= I - B\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1} + BCDA^{-1} \\
&\quad - BCDA^{-1}B\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1} \\
&= I - B\left[I+CDA^{-1}B\right]\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1} + BCDA^{-1} \\
&= I - BC\left[C^{-1}+DA^{-1}B\right]\left[C^{-1}+DA^{-1}B\right]^{-1}DA^{-1} + BCDA^{-1} \\
&= I - BCDA^{-1} + BCDA^{-1} = I.
\end{aligned}$$

□

When we study linear models, we frequently need to refer to matrices and vectors that consist entirely of 1s. Such matrices are denoted by the letter J with various subscripts and superscripts to specify their dimensions. J_r^c is an $r \times c$ matrix of 1s. The subscript indicates the number of rows and the superscript indicates the number of columns. If there is only one column, the superscript may be suppressed, e.g., $J_r = J_r^1$. In a context where we are dealing with vectors in $\mathbf{R}^n$, the subscript may also be suppressed, e.g., $J = J_n = J_n^1$.

A matrix of 0s is always denoted by 0.

B.5 Properties of Kronecker Products and Vec Operators

Kronecker products and Vec operators are extremely useful in multivariate analysis and some approaches to variance component estimation. They are also often used in writing balanced ANOVA models. We now present their basic algebraic properties.

1. If the matrices are of conformable sizes, $[A \otimes (B+C)] = [A \otimes B] + [A \otimes C]$.
2. If the matrices are of conformable sizes, $[(A+B) \otimes C] = [A \otimes C] + [B \otimes C]$.
3. If a and b are scalars, $ab[A \otimes B] = [aA \otimes bB]$.
4. If the matrices are of conformable sizes, $[A \otimes B][C \otimes D] = [AC \otimes BD]$.
5. The transpose of a Kronecker product matrix is $[A \otimes B]' = [A' \otimes B']$.
6. The generalized inverse of a Kronecker product matrix is $[A \otimes B]^- = [A^- \otimes B^-]$.
7. For two vectors v and w, $\mathrm{Vec}(vw') = w \otimes v$.
8. For a matrix W and conformable matrices A and B, $\mathrm{Vec}(AWB') = [B \otimes A]\mathrm{Vec}(W)$.

9. For conformable matrices A and B, $\text{Vec}(A)'\text{Vec}(B) = \text{tr}(A'B)$.
10. The Vec operator commutes with any matrix operation that is performed elementwise. For example, $\text{E}\{\text{Vec}(W)\} = \text{Vec}\{\text{E}(W)\}$ when W is a random matrix. Similarly, for conformable matrices A and B and scalar ϕ, $\text{Vec}(A+B) = \text{Vec}(A) + \text{Vec}(B)$ and $\text{Vec}(\phi A) = \phi\text{Vec}(A)$.
11. If A and B are positive definite, then $A \otimes B$ is positive definite.

Most of these are well-known facts and easy to establish. Two of them are somewhat more unusual, and we present proofs.

ITEM 8. We show that for a matrix W and conformable matrices A and B, $\text{Vec}(AWB') = [B \otimes A]\text{Vec}(W)$. First note that if $\text{Vec}(AW) = [I \otimes A]\text{Vec}(W)$ and $\text{Vec}(WB') = [B \otimes I]\text{Vec}(W)$, then $\text{Vec}(AWB') = [I \otimes A]\text{Vec}(WB') = [I \otimes A][B \otimes I]\text{Vec}(W) = [B \otimes A]\text{Vec}(W)$.

To see that $\text{Vec}(AW) = [I \otimes A]\text{Vec}(W)$, let W be $r \times s$ and write W in terms of its columns $W = [w_1, \ldots, w_s]$. Then $AW = [Aw_1, \ldots, Aw_s]$ and $\text{Vec}(AW)$ stacks the columns $Aw_1, \ldots, Aw_s$. On the other hand,

$$[I \otimes A]\text{Vec}(W) = \begin{bmatrix} A & & 0 \\ & \ddots & \\ 0 & & A \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_s \end{bmatrix} = \begin{bmatrix} Aw_1 \\ \vdots \\ Aw_s \end{bmatrix}.$$

To see that $\text{Vec}(WB') = [B \otimes I]\text{Vec}(W)$, take W as above and write $B_{m \times s} = [b_{ij}]$ with *rows* $b'_1, \ldots, b'_m$. First note that $WB' = [Wb_1, \ldots, Wb_m]$, so $\text{Vec}(WB')$ stacks the columns $Wb_1, \ldots, Wb_m$. Now observe that

$$[B \otimes I_r]\text{Vec}(W) = \begin{bmatrix} b_{11}I_r & \cdots & b_{1s}I_r \\ \vdots & \ddots & \vdots \\ b_{m1}I_r & \cdots & b_{ms}I_r \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_s \end{bmatrix} = \begin{bmatrix} Wb_1 \\ \vdots \\ Wb_m \end{bmatrix}.$$

ITEM 11. To see that if $A_{r \times r}$ and $B_{s \times s}$ are positive definite, then $A \otimes B$ is positive definite, consider the eigenvalues and eigenvectors of A and B. Recall that a symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Suppose that $Av = \phi v$ and $Bw = \theta w$. We now show that all of the eigenvalues of $A \otimes B$ are positive. Observe that

$$\begin{aligned} [A \otimes B][v \otimes w] &= [Av \otimes Bw] \\ &= [\phi v \otimes \theta w] \\ &= \phi\theta[v \otimes w]. \end{aligned}$$

This shows that $[v \otimes w]$ is an eigenvector of $[A \otimes B]$ corresponding to the eigenvalue $\phi\theta$. As there are r choices for ϕ and s choices for θ, this accounts for all rs of the eigenvalues in the $rs \times rs$ matrix $[A \otimes B]$. Moreover, ϕ and θ are both positive, so all of the eigenvalues of $[A \otimes B]$ are positive.

B.6 Tensors

Tensors are simply an alternative notation for writing vectors. This notation has substantial advantages when dealing with quadratic forms and when dealing with more general concepts than quadratic forms. Our main purpose in discussing them here is simply to illustrate how flexibly subscripts can be used in writing vectors.

Consider a vector $Y = (y_1, \ldots, y_n)'$. The tensor notation for this is simply y_i. We can write another vector $a = (a_1, \ldots, a_n)'$ as a_i. When written individually, the subscript is not important. In other words, a_i is the same vector as a_j. Note that the length of these vectors needs to be understood from the context. Just as when we write Y and a in conventional vector notation, there is nothing in the notation y_i or a_i to tell us how many elements are in the vector.

If we want the inner product $a'Y$, in tensor notation we write $a_i y_i$. Here we are using something called the *summation convention*. Because the subscripts on a_i and y_i are the same, $a_i y_i$ is taken to mean $\sum_{i=1}^n a_i y_i$. If, on the other hand, we wrote $a_i y_j$, this means something completely different. $a_i y_j$ is an alternative notation for the Kronecker product $[a \otimes Y] = (a_1 y_1, \ldots, a_1 y_n, a_2 y_1, \ldots, a_n y_n)'$. In $[a \otimes Y] \equiv a_i y_j$, we have two subscripts identifying the rows of the vector.

Now, suppose we want to look at a quadratic form $Y'AY$, where Y is an n vector and A is $n \times n$. One way to rewrite this is

$$Y'AY = \sum_{i=1}^n \sum_{j=1}^n y_i a_{ij} y_j = \sum_{i=1}^n \sum_{j=1}^n a_{ij} y_i y_j = \mathrm{Vec}(A)'[Y \otimes Y].$$

Here we have rewritten the quadratic form as a linear combination of the elements in the vector $[Y \otimes Y]$. The linear combination is determined by the elements of the vector $\mathrm{Vec}(A)$. In tensor notation, this becomes quite simple. Using the summation convention in which objects with the same subscript are summed over,

$$Y'AY = y_i a_{ij} y_j = a_{ij} y_i y_j.$$

The second term just has the summation signs removed, but the third term, which obviously gives the same sum as the second, is actually the tensor notation for $\mathrm{Vec}(A)'[Y \otimes Y]$. Again, $\mathrm{Vec}(A) = (a_{11}, a_{21}, a_{31}, \ldots, a_{nn})'$ uses two subscripts to identify rows of the vector. Obviously, if you had a need to consider things like

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} y_i y_j y_k \equiv a_{ijk} y_i y_j y_k,$$

the tensor version $a_{ijk} y_i y_j y_k$ saves some work.

There is one slight complication in how we have been writing things. Suppose A is not symmetric and we have another n vector W. Then we might want to consider

$$W'AY = \sum_{i=1}^n \sum_{j=1}^n w_i a_{ij} y_j.$$

From item 8 in the previous subsection,

$$W'AY = \mathrm{Vec}(W'AY) = [Y' \otimes W']\mathrm{Vec}(A).$$

Alternatively,

$$W'AY = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i a_{ij} y_j = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} y_j w_i = \mathrm{Vec}(A)'[Y \otimes W]$$

or $W'AY = Y'A'W = \mathrm{Vec}(A')'[W \otimes Y]$. However, with A nonsymmetric, $W'A'Y = \mathrm{Vec}(A')'[Y \otimes W]$ is typically different from $W'AY$. The Kronecker notation requires that care be taken in specifying the order of the vectors in the Kronecker product, and whether or not to transpose A before using the Vec operator. In tensor notation, $W'AY$ is simply $w_i a_{ij} y_j$. In fact, the orders of the vectors can be permuted in any way; so, for example, $a_{ij} y_j w_i$ means the same thing. $W'A'Y$ is simply $w_i a_{ji} y_j$. The tensor notation and the matrix notation require less effort than the Kronecker notation.

For our purposes, the real moral here is simply that the subscripting of an individual vector does not matter. We can write a vector $Y = (y_1, \ldots, y_n)'$ as $Y = [y_k]$ (in tensor notation as simply y_k), or we can write the same n vector as $Y = [y_{ij}]$ (in tensor notation, simply y_{ij}), where, as long as we know the possible values that i and j can take on, the actual order in which we list the elements is not of much importance. Thus, if $i = 1, \ldots, t$ and $j = 1, \ldots, N_i$, with $n = \sum_{i=1}^{t} N_i$, it really does not matter if we write a vector Y as $(y_1, \ldots, y_n)$, or $(y_{11}, \ldots, y_{1N_1}, y_{21}, \ldots, y_{tN_t})'$ or $(y_{t1}, \ldots, y_{tN_t}, y_{t-1,1}, \ldots, y_{1N_1})'$ or in any other fashion we may choose, as long as we keep straight which row of the vector is which. Thus, a linear combination $a'Y$ can be written $\sum_{k=1}^{n} a_k y_k$ or $\sum_{i=1}^{t}\sum_{j=1}^{N_i} a_{ij} y_{ij}$. In tensor notation, the first of these is simply $a_k y_k$ and the second is $a_{ij} y_{ij}$. These ideas become very handy in examining analysis of variance models, where the standard approach is to use multiple subscripts to identify the various observations. The subscripting has no intrinsic importance; the only thing that matters is knowing which row is which in the vectors. The subscripts are an aid in this identification, but they do not create any problems. We can still put all of the observations into a vector and use standard operations on them.

B.7 Exercises

Exercise B.1

(a) Show that

$$A^k x + b_{k-1} A^{k-1} x + \cdots + b_0 x = (A - \mu I)\left(A^{k-1} x + \tau_{k-2} A^{k-2} x + \cdots + \tau_0 x\right) = 0,$$

where μ is any nonzero solution of $b_0 + b_1 w + \cdots + b_k w^k = 0$ with $b_k = 1$ and $\tau_j = -(b_0 + b_1\mu + \cdots + b_j\mu^j)/\mu^{j+1}$, $j = 0, \ldots, k$.

(b) Show that if the only root of $b_0 + b_1 w + \cdots + b_k w^k$ is zero, then the factorization in (a) still holds.

(c) The solution μ used in (a) need not be a real number, in which case μ is a complex eigenvalue and the τ_is are complex; so the eigenvector is complex. Show that with A symmetric, μ must be real because the eigenvalues of A must be real. In particular, assume that

$$A(y + iz) = (\lambda + i\gamma)(y + iz),$$

for y, z, λ, and γ real vectors and scalars, respectively, set $Ay = \lambda y - \gamma z$, $Az = \gamma y + \lambda z$, and examine $z'Ay = y'Az$.

Exercise B.2 Prove Proposition B.32.

Exercise B.3 Show that any nonzero symmetric matrix A can be written as $A = PDP'$, where $C(A) = C(P)$, $P'P = I$, and D is nonsingular.

Exercise B.4 Prove Corollary B.52.

Exercise B.5 Prove Corollary B.53.

Exercise B.6 Show $\text{tr}(cI_n) = nc$.

Exercise B.7 Let a, b, c, and d be real numbers. If $ad - bc \neq 0$, find the inverse of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Exercise B.8 Prove Theorem B.28, i.e., let A be an $r \times s$ matrix, let B be an $s \times r$ matrix, and show that $\text{tr}(AB) = \text{tr}(BA)$.

Exercise B.9 Determine whether the matrices given below are positive definite, nonnegative definite, or neither.

$$\begin{bmatrix} 3 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 10 \end{bmatrix}, \quad \begin{bmatrix} 26 & -2 & -7 \\ -2 & 4 & -6 \\ -7 & -6 & 13 \end{bmatrix}, \quad \begin{bmatrix} 26 & 2 & 13 \\ 2 & 4 & 6 \\ 13 & 6 & 13 \end{bmatrix}, \quad \begin{bmatrix} 3 & 2 & -2 \\ 2 & -2 & -2 \\ -2 & -2 & 10 \end{bmatrix}.$$

Exercise B.10 Show that the matrix B given below is positive definite, and find a matrix Q such that $B = QQ'$. (Hint: The first row of Q can be taken as $(1,-1,0)$.)

$$B = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

Exercise B.11 Let

$$A = \begin{bmatrix} 2 & 0 & 4 \\ 1 & 5 & 7 \\ 1 & -5 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 5 & 1 \\ -3 & 0 & 1 \end{bmatrix}.$$

Use Theorem B.35 to find the perpendicular projection operator onto the column space of each matrix.

Exercise B.12 Show that for a perpendicular projection matrix M,

$$\sum_i \sum_j m_{ij}^2 = r(M).$$

Exercise B.13 Prove that if $M = M'M$, then $M = M'$ and $M = M^2$.

Exercise B.14 Let M_1 and M_2 be perpendicular projection matrices, and let M_0 be a perpendicular projection operator onto $C(M_1) \cap C(M_2)$. Show that the following are equivalent:

(a) $M_1 M_2 = M_2 M_1$.

(b) $M_1 M_2 = M_0$.

(c) $\left\{C(M_1) \cap [C(M_1) \cap C(M_2)]^\perp\right\} \perp \left\{C(M_2) \cap [C(M_1) \cap C(M_2)]^\perp\right\}$.

Hints: (i) Show that $M_1 M_2$ is a projection operator. (ii) Show that $M_1 M_2$ is symmetric. (iii) Note that $C(M_1) \cap [C(M_1) \cap C(M_2)]^\perp = C(M_1 - M_0)$.

Exercise B.15 Let M_1 and M_2 be perpendicular projection matrices. Show that

(a) the eigenvalues of $M_1 M_2$ have length no greater than 1 in absolute value (they may be complex);

(b) $\mathrm{tr}(M_1 M_2) \leq r(M_1 M_2)$.

Hints: For part (a) show that with $x'Mx \equiv \|Mx\|^2$, $\|Mx\| \leq \|x\|$ for any perpendicular projection operator M. Use this to show that if $M_1 M_2 x = \lambda x$, then $\|M_1 M_2 x\| \geq |\lambda| \, \|M_1 M_2 x\|$.

Exercise B.16 For vectors x and y, let $M_x = x(x'x)^{-1}x'$ and $M_y = y(y'y)^{-1}y'$. Show that $M_x M_y = M_y M_x$ if and only if $C(x) = C(y)$ or $x \perp y$.

Exercise B.17 Consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

(a) Show that A is a projection matrix.
(b) Is A a perpendicular projection matrix? Why or why not?
(c) Describe the space that A projects onto and the space that A projects along. Sketch these spaces.
(d) Find another projection operator onto the space that A projects onto.

Exercise B.18 Let A be an arbitrary projection matrix. Show that $C(I - A) = C(A')^{\perp}$.

Hints: Recall that $C(A')^{\perp}$ is the null space of A. Show that $(I - A)$ is a projection matrix.

Exercise B.19 Show that if A^- is a generalized inverse of A, then so is

$$G = A^-AA^- + (I - A^-A)B_1 + B_2(I - AA^-)$$

for any choices of B_1 and B_2 with conformable dimensions.

Exercise B.20 Let A be positive definite with eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that A^{-1} has eigenvalues $1/\lambda_1, \ldots, 1/\lambda_n$ and the same eigenvectors as A.

Exercise B.21 For A nonsingular, let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

and let $A_{1\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Show that if all inverses exist,

$$A^{-1} = \begin{bmatrix} A_{1\cdot 2}^{-1} & -A_{1\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{1\cdot 2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{1\cdot 2}^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

and that

$$A_{22}^{-1} + A_{22}^{-1}A_{21}A_{1\cdot 2}^{-1}A_{12}A_{22}^{-1} = \left[A_{22} - A_{21}A_{11}^{-1}A_{12}\right]^{-1}.$$