Matrix Algebra: Appendix A, B
Ronald Christensen

Hakbae Lee

The Department of Statistics and Data Science, Yonsei University

Definition A.1. A set $\mathcal{M} \subset \mathbf{R}^n$ is a vector space if, for any $x, y, z \in M$ and scalars α, β , operations of vector addition and scalar multiplication are defined such that:

(1) $(x + y) + z = x + (y + z)$.

(2) $x + y = y + x$.

(3) There exists a vector $0 \in \mathcal{M}$ such that $x + 0 = x = 0 + x$ for any $x \in \mathcal{M}$.

(4) For any $x \in \mathcal{M}$, there exists $y = -x$ such that $x + y = 0 = y + x$.

(5) $\alpha(x + y) = \alpha x + \alpha y$.

(6) $(\alpha + \beta)x = \alpha x + \beta x$.

(7) $(\alpha\beta)x = \alpha(\beta x)$.

(8) There exists a scalar $\xi = 1$ such that $\xi x = x$.

Definition A.2. Let $\mathcal{M}$ be a vector space, and let $\mathcal{N}$ be a set with $\mathcal{N} \subset \mathcal{M}$. $\mathcal{N}$ is a subspace of $\mathcal{M}$ if and only if $\mathcal{N}$ is a vector space.

Theorem A.3. Let $\mathcal{M}$ be a vector space, and let $\mathcal{N}$ be a nonempty subset of $\mathcal{M}$. If $\mathcal{N}$ is closed under vector addition and scalar multiplication, then $\mathcal{N}$ is a subspace of $\mathcal{M}$.

Theorem A.4. Let $\mathcal{M}$ be a vector space, and let $x_1, \ldots, x_r \in \mathcal{M}$. Then $\{v | v = \alpha_1 x_1 + \ldots + \alpha_r x_r, \alpha_i \in \mathbf{R}^n\} \subset \mathcal{M}$

Definition A.5. The set of all linear combinations of $x_1, \ldots, x_r$ is called the *space spanned* by $x_1, \ldots, x_r$. If $\mathcal{N}$ is a subspace of M, and $\mathcal{N}$ equals the space spanned by by $x_1, \ldots, x_r$, then $\{x_1, \ldots, x_r\}$ is called a spanning set for $\mathcal{N}$.

Consider an $n \times p$ matrix A

$$
\begin{aligned}
A &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ip} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix} = \begin{pmatrix} R_1^T \\ R_2^T \\ \vdots \\ R_n^T \end{pmatrix} \\
&= \begin{pmatrix} C_1 & C_2 & \dots & C_p \end{pmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
R_i^T &= i\text{-th } 1 \times p \text{ row vector, } i = 1, \dots, n \\
C_j &= j\text{-th } n \times 1 \text{ column vector, } j = 1, \dots, p \\
\mathcal{C} &= \text{Span}\{C_1, \dots, C_p\} = \text{column space} \subset \mathbf{R}^n \\
\mathcal{R} &= \text{Span}\{R_1, \dots, R_p\} = \text{row space} \subset \mathbf{R}^p
\end{aligned}
$$

Definition A.6. Let $x_1, \ldots, x_r \in \mathcal{M}$. If there exist scalars $\alpha_1, \ldots, \alpha_r$ not all zero so that $\sum_i \alpha_i x_1 = 0$, then $x_1, \ldots, x_r$ are linearly dependent. If such αs do not exist, $x_1, \ldots, x_r$ are linearly independent.

Definition A.7. If $\mathcal{N} \subset \mathcal{M}$ and if $\{x_1, ..., x_r\}$ is a linearly independent spanning set for $\mathcal{N}$, then $\{x_1, \ldots, x_r\}$ is called a basis for $\mathcal{N}$.

Theorem A.8. If $\mathcal{N} \subset \mathcal{M}$, all bases for $\mathcal{N}$ have the same number of vectors.

Theorem A.9. If $v_1, ..., v_r$ is a basis for $\mathcal{N}$, and $x \in \mathcal{N}$, then the characterization $x = \sum_{i=1}^{r} \alpha_1 v_i$ is unique.

Definition A.10.

$$r(\mathcal{N}) \equiv \text{rank of } \mathcal{N} = \text{No of vectors in a basis of } \mathcal{N}$$

Definition A.11.

- The (Euclidean) inner product between two vectors x and y in $\mathbf{R}^n$ is $x^T y$.
- Two vectors x and $y$are orthogonal, $x \perp y$ if $x^T y = 0$.
- Two subspaces $\mathcal{N}_1$ and $\mathcal{N}_2$ are orthogonal if $x \in \mathcal{N}_1$ and $y \in \mathcal{N}_2$ imply that $x^T y = 0$.
- $\{x_1, \ldots, x_r\}$ is an orthogonal basis for a space $\mathcal{N}$ if $\{x_1, \ldots, x_r\}$ is a basis for $\mathcal{N}$ and for $i \neq j$, $x_i^T x_j = 0$.
- $\{x_1, \ldots, x_r\}$ is an orthonormal basis for $\mathcal{N}$ if $\{x_1, \ldots, x_r\}$ is an orthogonal basis and $x_i^T x_1 = 1$ for $i = 1, \ldots r$.
- The terms orthogonal and perpendicular are used interchangeably.
- The length of a vector x is $||x|| \equiv \sqrt{x^T x}$. The distance between two vectors x and y is the length of their difference, i.e., $||x - y||$.

Definition A.12. The **Gram - Schmidt (Orthogonalization) Theorem.**
Let $\mathcal{N}$ be a space with basis $\{x_1, \ldots, x_r\}$. There exists an orthonormal basis for $\mathcal{N}$, say $\{y_1, \ldots, y_r\}$, with y_s in the space spanned by $\{x_1, \ldots, x_s, s = 1, \ldots r\}$.

Definition A.13. Let $\mathcal{N} \subset \mathcal{M}$.

$$\begin{aligned} \mathcal{N}^{\perp}_{\mathcal{M}} &\equiv \{y \in \mathcal{M} | y \perp \mathcal{N}\} \\ &= \text{orthogonal complement of } \mathcal{N} \text{ with respect to } \mathcal{M} \end{aligned}$$

Theorem A.14. Let $\mathcal{N} \subset \mathcal{M}$. Then $\mathcal{N}_{\mathcal{M}}^{\perp} \subset \mathcal{M}$: in other words, if $x \in \mathcal{M}$, x can be written uniquely as $x = x_0 + x_1$ with $x_0 \in \mathcal{N}$ and $x_1 \in \mathcal{N}_{\mathcal{M}}^{\perp}$ and $r(\mathcal{M}) = r(\mathcal{N}) + r(\mathcal{N}_{\mathcal{M}}^{\perp})$.

Definition A.15. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be vector subspaces. Then

$$\mathcal{N}_1 + \mathcal{N}_2 = \{x | x = x_1 + x_2, x_j \in \mathcal{N}_j, j = 1, 2\}$$

Theorem A.16. $\mathcal{N}_1 + \mathcal{N}_2$ is a vector space and $\mathcal{C}(A, B) = \mathcal{C}(A) + \mathcal{C}(B)$.

pp.415–416

Definition B.1. Square Matrix $p \times p$ matrix A

Definition B.2. Transpose of $A = [a_{ij}]$; $A^T = [a_{ji}]$

Definition B.3. Symmetric matrix; $A = A^T$

Definition B.4. Diagonal matrix of a square $A \equiv \text{diag}(\lambda_j)$

Definition B.5. Let A; $r \times c$ and B; $s \times d$.

$$A \otimes B \equiv \text{Kronecker product of } A \text{ and } B = \left(a_{ij}B\right) : rs \times cd$$

Definition B.6. Let $A = (A_1, \ldots, A_c)$; $r \times c$.

$$[\text{vec}(A)]^T = (A_1^T, A_2^T, \ldots, A_c^T)$$

Definition B.8. $A : n \times n$ is nonsingular if $A^{-1}A = AA^{-1} = I_n$ where A^{-1} is inverse of A

Theorem B.9 $A : n \times n$ is nonsingular if and only if $r(A) = n$.

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

Definition B.11.

$$\mathcal{N}(A) \equiv \{x | Ax = 0\} = \text{ Null space of } A$$

Theorem B.12. Let A be $n \times n$ and $r(A) = r$. Then $r(\mathcal{N}(A)) = n - r$.

Definition B.13. λ is an eigenvalue of A if $A - \lambda I$ is singular. λ is an eigenvalue of multiplicity s if $r = (\mathcal{N}(A - \lambda I)) = s$. A nonzero vector x is an eigenvector of A corresponding to λ if $Ax = \lambda x$. Eigenvalues are also called singular values and characteristic roots. ; pp. 421-422

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

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

Definition B.16. A square matrix P is orthogonal if $P^T = P^{-1}$.

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

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

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

proof: p.423

Spectral Decomposition(SD)
=Singular Value Decomposition(SVD) for a symmetric matrix

Corollary B.20.

$$A = PD(\lambda_i)P^T$$

Definition B.21. A symmetric matrix A is positive (nonnegative) definite, p.d(n.d) if, for any nonzero vector $v \in \mathbf{R}^n$, $v^T Av > 0 (\geq 0)$.

Theorem B.22. *A is nonnegative definite(n.d) if and only if there exists a square matrix Q such that $A = QQ^T$.*

PROOF; p.424

Corollary B.23. *A is positive definite(p.d) if and only if Q is nonsingular for any choice of Q.*

PROOF; p.424

Theorem B.24. If A is an $n \times n$ nonnegative definite matrix with nonzero eigenvalues $\lambda_1, \ldots, \lambda_r$, then there exists an $n \times r$ matrix $Q = Q_1 Q_2^{-1}$ such that Q_1 has orthonormal columns, $\mathcal{C}(Q1) = \mathcal{C}(A)$, Q_2 is diagonal and nonsingular, and $Q^T AQ = I_r$.

Corollary B.25. Let $W = Q_1 Q_2$. Then $W^T W = A$.

Corollary B.26. $AQQ^T A = A$ and $QQ^T AQQ^T = QQ^T$.

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 $B_{s\times r}$, $\mathrm{tr(AB) = tr(BA)}$.

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

Theorem B.30. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. then $\mathrm{tr(A) = \sum_{i=1}^{n} \lambda_i}$ and $\det(A) = \prod_{i=1}^{n} \lambda_1$.

Orthogonal(perpendicular) projection operator: opo (ppo)

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

(i) $v \in \mathcal{C}(X)$ implies $Mv = v$; projecton

(ii) $w \perp \mathcal{C}(X)$ implies $Mw = 0$; perpendicularity

Proposition B.32. If M is opo(ppo) onto $\mathcal{C}(X)$, then $\mathcal{C}(M) = \mathcal{C}(X)$.

Theorem B.33. M is opo(ppo) onto $\mathcal{C}(X)$ if and only if $MM = M$ and $M^T = M$.

PROOF; p.427

Proposition B.34. Perpendicular(orthogonal) projection operators are unique.

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

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}.

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

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

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

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

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

(i) $XGX^TX = XHX^TX = X$

(ii) $XGX^T = XHX^T$

Projections

Theorem B.44. $X(X^TX)^-X^T$ is the perpendicular projection operator onto $\mathcal{C}(X)$.

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 $\mathcal{C}(M_1, M_2)$ if and only if $\mathcal{C}(M_1) \perp \mathcal{C}(M_2)$.

Theorem B.46. Let M_1 and M_2 be symmetric. Then $\mathcal{C}(M_1) \perp \mathcal{C}(M_2)$, and $(M_1 + M_2)$ is a perpendicular projection matrix, then M_1 and M_2 are perpendicular projection matrices.

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

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

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

Definition B.50.

(a) If A is a square matrix with $A^2 = A$, then A is called idempotent.

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

Proposition B.51. For any matrix X, $\mathcal{C}(XX^T) = \mathcal{C}(X)$.

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

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

Proposition B.54. If B is nonsingular, $\mathcal{C}(XB) = \mathcal{C}(X)$.

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

Proposition B.56. When all inverses exist,

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}$$

Properties of Kronecker Products and Vec Operators

1. $[A \otimes (B + C)] = [A \otimes B] + [A \otimes C]$
2. $[(A + B) \otimes C] = [A \otimes C] + [B \otimes C]$
3. $ab[A \otimes B] = [aA \otimes bB]$ for $a, b \in \mathbf{R}$
4. $[A \otimes B][C \otimes D] = [AC \otimes BD]$
5. $[A \otimes B]^T = [A^T \otimes B^T]$
6. $[A \otimes B]^- = [A^- \otimes B^-]$
7. $\text{Vec}(vw^T) = w \otimes v$ for two vectors v, w
8. $\text{Vec}(AWB^T) = [B \otimes A]\text{Vec}(W)$
9. $\text{Vec}(A)^T\text{Vec}(B) = \mathrm{tr}(\mathrm{A}^\mathrm{T}\mathrm{B})$
10. $\text{E}\{\text{Vec}(W)\} = \text{Vec}\{\text{E}(W)\}$ for a random matrix W
    $\text{Vec}(A + B) = \text{Vec}(A) + \text{Vec}(B)$
    $\text{Vec}(\phi A) = \phi\text{Vec}(A)$ for $\phi \in \mathbf{R}$
11. Let A and B be positive definite. Then $A \otimes B$ is positive definite.