Comments on Selected Problems

CHAPTER 1

4. This problem gives the direct sum version of partitioned matrices. For (ii), identify V_1 with vectors of the form $\{v_1, 0\} \in V_1 \oplus V_2$ and restrict T to these. This restriction is a map from V_1 to $V_1 \oplus V_2$ so $T\{v_1, 0\} = \{z_1(v_1), z_2(v_1)\}$ where $z_1(v_1) \in V_1$ and $z_2(v_1) \in V_2$. Show that z_1 is a linear transformation on V_1 to V_1 and z_2 is a linear transformation on V_1 to V_2. This gives A_{11} and A_{21}. A similar argument gives A_{12} and A_{22}. Part (iii) is a routine computation.
5. If $x_{r+1} = \sum_1^r c_i x_i$, then $w_{r+1} = \sum_1^r c_i w_i$.
8. If $u \in R^k$ has coordinates $u_1, \ldots, u_k$, then $Au = \sum_1^k u_i x_i$ and all such vectors are just span $\{x_1, \ldots, x_k\}$. For (ii), $r(A) = r(A')$ so $\dim \mathfrak{R}(A'A) = \dim \mathfrak{R}(AA')$.
10. The algorithm of projecting $x_2, \ldots, x_k$ onto $\{\text{span } x_1\}^{\perp}$ is known as Björk's algorithm (Björk, 1967) and is an alternative method of doing Gram–Schmidt. Once you see that $y_2, \ldots, y_k$ are perpendicular to y_1, this problem is not hard.
11. The assumptions and linearity imply that $[Ax, w] = [Bx, w]$ for all $x \in V$ and $w \in W$. Thus $[(A - B)x, w] = 0$ for all w. Choose $w = (A - B)x$ so $(A - B)x = 0$.
12. Choose z such that $[y_1, z] \neq 0$. Then $[y_1, z]x_1 = [y_2, z]x_2$ so set $c = [y_2, z]/[y_1, z]$. Thus $cx_2 \square y_1 = x_2 \square y_2$ so $cy_1 \square x_2 = y_2 \square x_2$. Hence $c\|x_2\|^2 y_1 = \|x_2\|^2 y_2$ so $y_1 = c^{-1} y_2$.
13. This problem shows the topologies generated by inner products are all the same. We know $[x, y] = (x, Ay)$ for some $A > 0$. Let c_1 be the minimum eigenvalue of A, and let c_2 be the maximum eigenvalue of A.

14. This is just the Cauchy–Schwarz Inequality.
15. The classical two-way *ANOVA* table is a consequence of this problem. That A, B_1, B_2, and B_3 are orthogonal projections is a routine but useful calculation. Just keep the notation straight and verify that $P^2 = P = P'$, which characterizes orthogonal projections.
16. To show that $\Gamma(M^{\perp}) \subseteq M^{\perp}$, verify that $(u, \Gamma v) = 0$ for all $u \in M$ when $v \in M^{\perp}$. Use the fact that $\Gamma'\Gamma = I$ and $u = \Gamma u_1$ for some $u_1 \in M$ (since $\Gamma(M) \subseteq M$ and Γ is nonsingular).
17. Use Cauchy–Schwarz and the fact that $P_M x = x$ for $x \in M$.
18. This is Cauchy–Schwarz for the non-negative definite bilinear form $[C, D] = \operatorname{tr} ACBD'$.
20. Use Proposition 1.36 and the assumption that A is real.
21. The representation $\alpha P + \beta(I - P)$ is a spectral type representation—see Theorem 1.2a. If $M = \mathfrak{R}(P)$, let $x_1, \ldots, x_r, x_{r+1}, \ldots, x_n$ be any orthonormal basis such that $M = \operatorname{span}\{x_1, \ldots, x_r\}$. Then $Ax_i = \alpha x_i$, $i = 1, \ldots, r$, and $Ax_i = \beta x_i$, $i = r+1, \ldots, n$. The characteristic polynomial of A must be $(\alpha - \lambda)^r(\beta - \lambda)^{n-r}$.
22. Since $\lambda_1 = \sup_{\|x\|=1}(x, Ax)$, $\mu_1 = \sup_{\|x\|=1}(x, Bx)$, and $(x, Ax) \geqslant (x, Bx)$, obviously $\lambda_1 \geqslant \mu_1$. Now, argue by contradiction—let j be the smallest index such that $\lambda_j < \mu_j$. Consider eigenvectors $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ with $Ax_i = \lambda_i x_i$ and $By_i = \mu_i y_i$, $i = 1, \ldots, n$. Let $M = \operatorname{span}\{x_j, x_{j+1}, \ldots, x_n\}$ and let $N = \operatorname{span}\{y_1, \ldots, y_j\}$. Since dim $M = n - j + 1$, dim $M \cap N \geqslant 1$. Using the identities $\lambda_j = \sup_{x \in M, \|x\|=1}(x, Ax)$, $\mu_j = \inf_{x \in N, \|x\|=1}(x, Bx)$, for any $x \in M \cap N$, $\|x\| = 1$, we have $(x, Ax) \leqslant \lambda_j < \mu_j \leqslant (x, Bx)$, which is a contradiction.
23. Write $S = \sum_1^n \lambda_i x_i \Box x_i$ in spectral form where $\lambda_i > 0$, $i = 1, \ldots, n$. Then $0 = \langle S, T \rangle = \sum_1^n \lambda_i (x_i, Tx_i)$, which implies $(x_i, Tx_i) = 0$ for $i = 1, \ldots, n$ as $T \geqslant 0$. This implies $T = 0$.
24. Since $\operatorname{tr} A$ and $\langle A, I \rangle$ are both linear in A, it suffices to show equality for A's of the form $A = x \Box y$. But $\langle x \Box y, I \rangle = (x, y)$. However, that $\operatorname{tr} x \Box y = (x, y)$ is easily verified by choosing a coordinate system.
25. Parts (i) and (ii) are easy but (iii) is not. It is false that $A^2 \geqslant B^2$ and a 2×2 matrix counter example is not hard to construct. It is true that $A^{1/2} \geqslant B^{1/2}$. To see this, let $C = B^{1/2}A^{-1/2}$, so by hypothesis, $I \geqslant C'C$. Note that the eigenvalues of C are real and positive—being the same as those of $B^{1/4}A^{-1/2}B^{1/4}$ which is positive definite. If λ is any eigenvalue for C, there is a corresponding eigenvector—say x such that $\|x\| = 1$ and $Cx = \lambda x$. The relation $I \geqslant C'C$ implies $\lambda^2 \leqslant 1$, so $0 < \lambda \leqslant 1$ as λ is positive. Thus all the eigenvalues of C are in $(0, 1]$ so

the same is true of $A^{-1/4}B^{1/2}A^{-1/4}$. Hence $A^{-1/4}B^{1/2}A^{-1/4} \leqslant I$ so $B^{1/2} \leqslant A^{1/2}$.

26. Since P is an orthogonal projection, all its eigenvalues are zero or one and the multiplicity of one is the rank of P. But $\operatorname{tr} P$ is just the sum of the eigenvalues of P.

28. Since any $A \in \mathcal{L}(V, V)$ can be written as $(A + A')/2 + (A - A')/2$, it follows that $M + N = \mathcal{L}(V, V)$. If $A \in M \cap N$, then $A = A' = -A$, so $A = 0$. Thus $\mathcal{L}(V, V)$ is the direct sum of M and N so $\dim M + \dim N = n^2$. A direct calculation shows that $\{x_i \Box x_j + x_j \Box x_i | i \leqslant j\} \cup \{x_i \Box x_j - x_j \Box x_i | i < j\}$ is an orthogonal set of vectors, none of which is zero, and hence the set is linearly independent. Since the set has n^2 elements, it forms a basis for $\mathcal{L}(V, V)$. Because $x_i \Box x_j + x_j \Box x_i \in M$ and $x_i \Box x_j - x_j \Box x_i \in N$, $\dim M \geqslant n(n+1)/2$ and $\dim N \geqslant n(n-1)/2$. Assertions (i), (ii), and (iii) now follow. For (iv), just verify that the map $A \to (A + A')/2$ is idempotent and self-adjoint.

29. Part (i) is a consequence of $\sup_{\|v\|=1}\|Av\| = \sup_{\|v\|=1}[Av, Av]^{1/2} = \sup_{\|v\|=1}(v, A'Av)^{1/2}$ and the spectral theorem. The triangle inequality follows from $|||A + B||| = \sup_{\|v\|=1} \|Av + Bv\| \leqslant \sup_{\|v\|=1}(\|Av\| + \|Bv\|) \leqslant \sup_{\|v\|=1} \|Av\| + \sup_{\|v\|=1} \|Bv\|$.

30. This problem is easy, but it is worth some careful thought—it provides more evidence that $A \otimes B$ has been defined properly and $\langle \cdot, \cdot \rangle$ is an appropriate inner produce on $\mathcal{L}(W, V)$. Assertion (i) is easy since $(A \otimes B)(x_i \Box w_j) = (Ax_i) \Box (Bw_j) = (\lambda_i x_i) \Box (\mu_j w_j) = \lambda_i \mu_j x_i \Box w_j$. Obviously, $x_i \Box w_j$ is an eigenvector of the eigenvalue $\lambda_i \mu_j$. Part (ii) follows since the two linear transformations agree on the basis $\{x_i \Box w_j | i = 1, \ldots, m,\ j = 1, \ldots, n\}$ for $\mathcal{L}(W, V)$. For (iii), if the eigenvalues of A and B are positive, so are the eigenvalues of $A \otimes B$. Since the trace of a self-adjoint linear transformation in the sum of the eigenvalues (this is true even without self-adjointness, but the proof requires a bit more than we have established here), we have $\operatorname{tr} A \otimes B = \Sigma_{i,j}\lambda_i\mu_j = (\Sigma_i\lambda_i)(\Sigma_j\mu_j) = (\operatorname{tr} A)(\operatorname{tr} B)$. Since the determinant is the product of the eigenvalues, $\det(A \otimes B) = \Pi_{i,j}(\lambda_i\mu_j) = (\Pi\lambda_i)^n(\Pi\mu_j)^m = (\det A)^n(\det B)^m$.

31. Since $\psi'\psi = I_p$, ψ is a linearly isometry and its columns form an orthonormal set. Since $R(\psi) \subseteq M$ and the two subspaces have the same dimension, (i) follows. (ii) is immediate.

32. If C is $n \times k$ and D is $k \times n$, the set of nonzero eigenvalues of CD is the same as the set of nonzero eigenvalues of DC.

33. Apply Problem 32.

34. Orthogonal transformations preserve angles.

35. This problem requires that you have a facility in dealing with conditional expectation. If you do, the problem requires a bit of calculation but not much more. If you don't, proceed to Chapter 2.

CHAPTER 2

1. Write $x = \Sigma_1^n c_i x_i$ so $(x, X) = \Sigma c_i(x_i, X)$. Thus $\mathcal{E}|(x, X)| \leqslant \Sigma_1^n |c_i| \mathcal{E}|(x_i, X)|$ and $\mathcal{E}|(x_i, X)|$ is finite by assumption. To show that $\operatorname{Cov}(X)$ exists, it suffices to verify that $\operatorname{var}(x, X)$ exists for each $x \in V$. But $\operatorname{var}(x, X) = \operatorname{var}\{\Sigma c_i(x_i, X)\} = \Sigma\Sigma \operatorname{cov}\{c_i(x_i, X), c_j(x_j, X)\}$. Then $\operatorname{var}\{c_i(x_i, X)\} = \mathcal{E}[c_i(x_i X)]^2 - [\mathcal{E}c_i(x_i, X)]^2$, which exists by assumption. The Cauchy–Schwarz Inequality shows that $[\operatorname{cov}\{c_i(x_i, X), c_j(x_j, X)\}]^2 \leqslant \operatorname{var}\{c_i(x_i, X)\} \operatorname{var}\{c_j(x_j, X)\}$. But, $\operatorname{var}\{c_i(x_i, X)\}$ exists by the above argument.
2. All inner products on a finite dimensional vector space are related via the positive definite quadratic forms. An easy calculation yields the result of this problem.
3. Let $(\cdot, \cdot)_i$ be an inner product on V_i, $i = 1, 2$. Since f_i is linear on V_i, $f_i(x) = (x_i, x)_i$ for $x_i \in V_i$, $i = 1, 2$. Thus if X_1 and X_2 are uncorrelated (the choice of inner product is irrelevant by Problem 2), (2.2) holds. Conversely, if (2.2) holds, then $\operatorname{Cov}\{(x_1, X_1)_1, (x_2, X_2)_2\} = 0$ for $x_i \in V_i$, $i = 1, 2$ since $(x_1, \cdot)_1$ and $(x_2, \cdot)_2$ are linear functions.
4. Let $s = n - r$ and consider $\Gamma \in \mathcal{O}_r$ and a Borel set B_1 of R^r. Then

$$\begin{aligned} Pr\{\Gamma\dot{X} \in B_1\} &= Pr\{\Gamma\dot{X} \in B_1, \ddot{X} \in R^s\} \\ &= Pr\left\{\begin{pmatrix} \Gamma & 0 \\ 0 & I_s \end{pmatrix}\begin{pmatrix} \dot{X} \\ \ddot{X} \end{pmatrix} \in B_1 \times R^s\right\} \\ &= Pr\left\{\begin{pmatrix} \dot{X} \\ \ddot{X} \end{pmatrix} \in B_1 \times R^s\right\} = Pr\{\dot{X} \in B_1\}. \end{aligned}$$

The third equality holds since the matrix

$$\begin{pmatrix} \Gamma & 0 \\ 0 & I_s \end{pmatrix}$$

is in $\mathcal{O}_n$. Thus $\dot{X}$ has an $\mathcal{O}_r$-invariant distribution. That $\dot{X}$ given $\ddot{X}$ has an $\mathcal{O}_r$-invariant distribution is easy to prove when X has a density with respect to Lebesgue measure on R^n (the density has a version that satisfies $f(x) = f(\psi x)$ for $x \in R^n$, $\psi \in \mathcal{O}_n$). The general case requires some fiddling with conditional expections—this is left to the interested reader.

5. Let $A_i = \text{Cov}(X_i)$, $i = 1, \dots, n$. It suffices to show that $\text{var}(x, \Sigma X_i) = \Sigma(x, A_i x)$. But (x, X_i), $i = 1, \dots, n$, are uncorrelated, so $\text{var}[\Sigma(x, X_i)] = \Sigma\,\text{var}(x, X_i) = \Sigma(x, A_i x)$.
6. $\mathcal{E}U = \Sigma p_i \varepsilon_i = p$. Let U have coordinates $U_1, \dots, U_k$. Then $\text{Cov}(U) = \mathcal{E}UU' - pp'$ and UU' is a $p \times p$ matrix with elements U_iU_j. For $i \neq j$, $U_iU_j = 0$ and for $i = j$, $U_iU_j = U_i$. Since $\mathcal{E}U_i = p_i$, $\mathcal{E}UU' = D_p$. When $0 < p_i < 1$, D_p has rank k and the rank of $\text{Cov}(U)$ is the rank of $I_k - D_p^{-1/2}pp'D_p^{-1/2}$. Let $u = D_p^{-1/2}p$, so $u \in R^k$ has length one. Thus $I_k - uu'$ is a rank $k - 1$ orthogonal projection. The null space of $\text{Cov}\,U$ is $\text{span}\{e\}$ where e is the vector of ones in R^k. The rest is easy.
7. The random variable X takes on $n!$ values—namely the $n!$ permutations of x—each with probability $1/n!$. A direct calculation gives $\mathcal{E}X = \bar{x}e$ where $\bar{x} = n^{-1}\Sigma_1^n x_i$. The distribution of X is permutation invariant, which implies that $\text{Cov}\,X$ has the form $\sigma^2 A$ where $a_{ii} = 1$ and $a_{ij} = \rho$ for $i \neq j$ where $-1/(n-1) \leqslant \rho \leqslant 1$. Since $\text{var}(e'X) = 0$, we see that $\rho = -1/(n-1)$. Thus $\sigma^2 = \text{var}(X_1) = n^{-1}[\Sigma_1^n(x_i - \bar{x})^2]$ where X_1 is the first coordinate of X.
8. Setting $D = -I$, $\mathcal{E}X = -\mathcal{E}X$ so $\mathcal{E}X = 0$. For $i \neq j$, $\text{cov}\{X_i, X_j\} = \text{cov}\{-X_i, X_j\} = -\text{cov}\{X_i, X_j\}$ so X_i and X_j are uncorrelated. The first equality is obtained by choosing D with $d_{ii} = -1$ and $d_{jj} = 1$ in the relation $\mathcal{L}(X) = \mathcal{L}(DX)$.
9. This is a direct calculation.
10. It suffices to verify the equality for $A = x \,\square\, y$ as both sides of the equality are linear in A. For $A = x \,\square\, y$, $\langle A, \Sigma\rangle = (x, \Sigma y)$ and $(\mu, A\mu) = (\mu, x)(\mu, y)$, so the equality is obvious.
11. To say $\text{Cov}(X) = I_n \otimes \Sigma$ is to say that $\text{cov}\{(\text{tr}\,AX'), (\text{tr}\,BX')\} = \text{tr}\,A\Sigma B'$. To show rows 1 and 2 are uncorrelated, pick $A = \varepsilon_1 v'$ and $B = \varepsilon_2 u'$ where $u, v \in R^p$. Let X_1' and X_2' be the first two rows of X. Then $\text{tr}\,AX' = v'X_1$, $\text{tr}\,BX' = u'X_2$, and $\text{tr}\,A\Sigma B = 0$. The desired equality is established by first showing that it is valid for $A = xy'$, $x, y \in R^n$, and using linearity. When $A = xy'$, a useful equality is $X'AX = \Sigma_i\Sigma_j x_i y_j X_i X_j'$ where the rows of X are $X_1', \dots, X_n'$.
12. The equation $\Gamma A \Gamma' = A$ for $\Gamma \in \mathcal{O}_p$ implies that $A = cI_p$ for some c.
13. $\text{Cov}((\Gamma \otimes I)X) = \text{Cov}(X)$ implies $\text{Cov}(X) = I \otimes \Sigma$ for some Σ. $\text{Cov}((I \otimes \psi)X) = \text{Cov}(X)$ then implies $\psi\Sigma\psi' = \Sigma$, which necessitates $\Sigma = cI$ for some $c \geqslant 0$. Part (ii) is immediate since $\Gamma \otimes \psi$ is an orthogonal transformation on $(\mathcal{L}(V, W), \langle \cdot, \cdot \rangle)$.

14. This problem is a nasty calculation intended to inspire an appreciation for the equation $\mathrm{Cov}(X) = I_n \otimes \Sigma$.
15. Since $\mathcal{L}(X) = \mathcal{L}(-X)$, $\mathcal{E}X = 0$. Also, $\mathcal{L}(X) = \mathcal{L}(\Gamma X)$ implies $\mathrm{Cov}(X) = cI$ for some $c > 0$. But $\|X\|^2 = 1$ implies $c = 1/n$. Best affine predictor of X_1 given $\dot{X}$ is 0. I would predict X_1 by saying that X_1 is $\sqrt{1 - \dot{X}'\dot{X}}$ with probability $\frac{1}{2}$ and X_1 is $-\sqrt{1 - \dot{X}'\dot{X}}$ with probability $\frac{1}{2}$.
16. This is just the definition of $\square$.
17. For (i), just calculate. For (ii), $\mathrm{Cov}(S) = 2I_2 \otimes I_2$ by Proposition 2.23. The coordinate inner product on R^3 is not the inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{S}_2$.

CHAPTER 3

2. Since $\mathrm{var}(X_1) = \mathrm{var}(Y_1) = 1$ and $\mathrm{cov}\{X_1, Y_1\} = \rho$, $|\rho| \leqslant 1$. Form $Z = (XY)$—an $n \times 2$ matrix. Then $\mathrm{Cov}(Z) = I_n \otimes A$ where

$$A = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

When $|\rho| < 1$, A is positive definite, so $I_n \otimes A$ is positive definite. Conditioning on Y, $\mathcal{L}(X|Y) = N(\rho Y, (1 - \rho^2)I_n)$, so $\mathcal{L}(Q(Y)X|Y) = N(0, (1 - \rho^2)Q(Y))$ as $Q(Y)Y = 0$ and $Q(Y)$ is an orthogonal projection. Now, apply Proposition 3.8 for Y fixed to get $\mathcal{L}(W) = (1 - \rho^2)\chi^2_{n-1}$.
3. Just do the calculations.
4. Since $p(x)$ is zero in the second and fourth quadrants, X cannot be normal. Just find the marginal density of X_1 to show that X_1 is normal.
5. Write U in the form $X'AX$ where A is symmetric. Then apply Propositions 3.8 and 3.11.
6. Note that $\mathrm{Cov}(X \square X) = 2I \otimes I$ by Proposition 2.23. Since $(X, AX) = \langle X \square X, A \rangle$, and similarly for (X, BX), $0 = \mathrm{cov}\{(X, AX), (X, BX)\} = \mathrm{cov}\{\langle X \square X, A\rangle, \langle X \square X, B \rangle\} = \langle A, 2(I \otimes I)B\rangle = 2\,\mathrm{tr}\, AB$. Thus $0 = \mathrm{tr}\, A^{1/2}BA^{1/2}$ so $A^{1/2}BA^{1/2} = 0$, which shows $A^{1/2}B^{1/2} = 0$ and hence $AB = 0$.
7. Since $\mathcal{E}[\exp(itW_j)] = \exp\{it\mu_j - \sigma_j|t|]$, $\mathcal{E}[\exp(it\Sigma a_j W_j)] = \exp[it\Sigma a_j \mu_j - (\Sigma|a_j|\sigma_j)|t|]$, so $\mathcal{L}(\Sigma a_j W_j) = C(\Sigma a_j \mu_j, \Sigma|a_j|\sigma_j)$. Part (ii) is immediate from (i).
8. For (i), use the independence of R and Z_0 to compute as follows: $P\{U \leqslant u\} = P\{Z_0 \leqslant u/R\} = \int_0^\infty P\{Z_0 \leqslant u/t\}G(dt) = \int_0^\infty \Phi(u/t) G(dt)$ where Φ is the distribution function of Z_0. Now, differentiate. Part (ii) is clear.

9. Let $\mathfrak{B}_1$ be the sub σ-algebra induced by $T_1(X) = X_2$ and let $\mathfrak{B}_2$ be the sub σ-algebra induced by $T_2(X) = X_2'X_2$. Since $\mathfrak{B}_2 \subseteq \mathfrak{B}_1$, for any bounded function $f(X)$, we have $\mathcal{E}(f(X)|\mathfrak{B}_2) = \mathcal{E}(\mathcal{E}(f(X)|\mathfrak{B}_1)|\mathfrak{B}_2)$. But for $f(X) = h(X_2'X_1)$, the conditional expectation given $\mathfrak{B}_1$ can be computed via the conditional distribution of $X_2'X_1$ given X_2, which is

$$(3.3) \qquad \mathcal{L}(X_2'X_1|X_2) = N\left(X_2'X_2\Sigma_{22}^{-1}\Sigma_{21}, X_2'X_2 \otimes \Sigma_{11\cdot 2}\right).$$

Hence $\mathcal{E}(h(X_2'X_1)\mathfrak{B}_1)$ is $\mathfrak{B}_2$ measurable, so $\mathcal{E}(h(X_2'X_1)|\mathfrak{B}_2) = \mathcal{E}(h(X_2'X_1)|\mathfrak{B}_1)$. This implies that the conditional distribution (3.3) serves as a version of the conditional distribution of $X_2'X_1$ given $X_2'X_2$.

10. Show that $T^{-1}T_1 : R^n \to R^n$ is an orthogonal transformation so $l(C) = l((T^{-1}T_1)(C))$. Setting $B = T_1(C)$, we have $\nu_0(B) = \nu_1(B)$ for Borel B.

11. The measures ν_0 and ν_1 are equal up to a constant so all that needs to be calculated is $\nu_0(C)/\nu_1(C)$ for some set C with $0 < \nu_1(C) < +\infty$. Do the calculation for $C = \{v|[v, v] \leqslant 1\}$.

12. The inner product $\langle \cdot , \cdot \rangle$ on $\mathbb{S}_p$ is not the coordinate inner product. The "Lebesgue measure" on $(\mathbb{S}_p, \langle \cdot , \cdot \rangle)$ given by our construction is not $l(dS) = \prod_{i\leqslant j} ds_{ij}$, but is $\nu_0(dS) = (\sqrt{2})^{p(p-1)}l(dS)$.

13. Any matrix M of the form

$$M = a\begin{pmatrix} 1 & b & \cdots & b \\ b & 1 & & \vdots \\ \vdots & & \ddots & b \\ b & \cdots & b & 1 \end{pmatrix} : p \times p$$

can be written as $M = a[(p-1)b + 1]A + a(1-b)(I - A)$. This is a spectral decomposition for M so M has eigenvalues $a((p-1)b+1)$ and $a(1-b)$ (of multiplicity $p-1$). Setting $\alpha = a[(p-1)b+1]$ and $\beta = a(1-b)$ solves (i). Clearly, $M^{-1} = \alpha^{-1}A + \beta^{-1}(I - A)$ whenever α and β are not zero. To do part (ii), use the parameterization (μ, α, β) given above ($a = \sigma^2$ and $b = \rho$). Then use the factorization criterion on the likelihood function.

CHAPTER 4

1. Part (i) is clear since $Z\beta = \sum_1^k \beta_i z_i$ for $\beta \in R^k$. For (ii), use the singular value decomposition to write $Z = \sum_1^r \lambda_i x_i u_i'$ where r is the rank of Z, $\{x_1, \ldots, x_r\}$ is an orthonormal set in R^n, $\{u_1, \ldots, u_r\}$ is an orthonormal set in R^k, $M = \text{span}\{x_1, \ldots, x_r\}$, and $\mathfrak{N}(Z) = (\text{span}\{u_1, \ldots, u_r\})^{\perp}$.

Thus $(Z'Z)^- = \Sigma_1^r \lambda_i^{-2} u_i u_i'$ and a direct calculation shows that $Z(Z'Z)^- Z' = \Sigma_1^r x_i x_i'$, which is the orthogonal projection onto M.

2. Since $\mathcal{L}(X_i) = \mathcal{L}(\beta + \varepsilon_i)$ where $\mathcal{E}\varepsilon_i = 0$ and $\operatorname{var}(\varepsilon_i) = 1$, it follows that $\mathcal{L}(X) = \mathcal{L}(\beta e + \varepsilon)$ where $\mathcal{E}\varepsilon = 0$ and $\operatorname{Cov}(\varepsilon) = I_n$. A direct application of least-squares yields $\hat{\beta} = \bar{X}$ for this linear model. For (iii), since the same β is added to each coordinate of ε, the vector of ordered X's has the same distribution as the $\beta e + \nu$ where ν is the vector of ordered ε's. Thus $\mathcal{L}(U) = \mathcal{L}(\beta e + \nu)$ so $\mathcal{E}U = \beta e + a_0$ and $\operatorname{Cov}(U) = \operatorname{Cov}(\nu) = \Sigma_0$. Hence $\mathcal{L}(U - a_0) = \mathcal{L}(\beta e + (\nu - a_0))$. Based on this model, the Gauss–Markov estimator for β is $\tilde{\beta} = (e'\Sigma_0^{-1}e)^{-1}e'\Sigma_0^{-1}(U - a_0)$. Since $\bar{X} = (1/n)e'(U - a_0)$ (show $e'a_0 = 0$ using the symmetry of f), it follows from the Gauss–Markov Theorem that $\operatorname{var}(\tilde{\beta}) < \operatorname{var}(\hat{\beta})$.

3. That $M - \omega = M \cap \omega^{\perp}$ is clear since $\omega \subseteq M$. The condition $(P_M - P_\omega)^2 = P_M - P_\omega$ follows from observing that $P_M P_\omega = P_\omega P_M = P_\omega$. Thus $P_M - P_\omega$ is an orthogonal projection onto its range. That $\mathcal{R}(P_M - P_\omega) = M - \omega$ is easily verified by writing $x \in V$ as $x = x_1 + x_2 + x_3$ where $x_1 \in \omega$, $x_2 \in M - \omega$, and $x_3 \in M^{\perp}$. Then $(P_M - P_\omega)(x_1 + x_2 + x_3) = x_1 + x_2 - x_1 = x_2$. Writing $P_M = P_M - P_\omega + P_\omega$ and noting that $(P_M - P_\omega)P_\omega = 0$ yields the final identity.

4. That $\mathcal{R}(A) = M_0$ is clear. To show $\mathcal{R}(B_1) = M_1 - M_0$, first consider the transformation C defined by $(Cy)_{ij} = \bar{y}_{i\cdot}$, $i = 1, \dots, I$, $j = 1, \dots, J$. Then $C^2 = C = C'$, and clearly, $\mathcal{R}(C) \subseteq M_1$. But if $y \in M_1$, then $Cy = y$ so C is the orthogonal projection onto M_1. From Problem 3 (with $M = M_1$ and $\omega = M_0$), we see that $C - A_0$ is the orthogonal projection onto $M_1 - M_0$. But $((C - A_0)y)_{ij} = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$, which is just $(B_1 y)_{ij}$. Thus $B_1 = C - A_0$ so $\mathcal{R}(B_1) = M_1 - M_0$. A similar argument shows $\mathcal{R}(B_2) = M_2 - M_0$. For (ii), use the fact that $A_0 + B_1 + B_2 + B_3$ is the identity and the four orthogonal projections are perpendicular to each other. For (iii), first observe that $M = M_1 + M_2$ and $M_1 \cap M_2 = M_0$. If μ has the assumed representation, let ν be the vector with $\nu_{ij} = \alpha + \beta_i$ and let ξ be the vector with $\xi_{ij} = \gamma_j$. Then $\nu \in M_1$ and $\xi \in M_2$ so $\mu = \nu + \xi \in M_1 + M_2$.. Conversely, suppose $\mu \in M_0 \oplus (M_1 - M_0) \oplus (M_2 - M_0)$—say $\mu = \delta + \nu + \xi$. Since $\delta \in M_0$, $\delta_{ij} = \bar{\delta}_{\cdot\cdot}$ for all i, j, so set $\alpha = \bar{\delta}_{\cdot\cdot}$. Since $\nu \in M_1 - M_0$, $\nu_{ij} - \nu_{ik} = 0$ for all j, k for each fixed i and $\bar{\nu}_{\cdot\cdot} = 0$. Take $j = 1$ and set $\beta_i = \nu_{i1}$. Then $\nu_{ij} = \beta_i$ for $j = 1, \dots, J$ and, since $\bar{\nu}_{\cdot\cdot} = 0$, $\Sigma\beta_i = 0$. Similarly, setting $\gamma_j = \xi_{1j}$, $\xi_{ij} = \gamma_j$ for all i, j and since $\bar{\xi}_{\cdot\cdot} = 0$, $\Sigma\gamma_j = 0$. Thus $\mu_{ij} = \alpha + \beta_i + \gamma_j$ where $\Sigma\beta_i = \Sigma\gamma_j = 0$.

5. With $n = \dim V$, the density of Y is (up to constants) $f(y|\mu, \sigma^2) = \sigma^{-n}\exp[-(1/2\sigma^2)\|y - \mu\|^2]$. Using the results and notation Problem

3, write $V = \omega \oplus (M - \omega) \oplus M^{\perp}$ so $(M - \omega) \oplus M^{\perp} = \omega^{\perp}$. Under H_0, $\mu \in \omega$ so $\hat{\mu}_0 = P_\omega y$ is the maximum likelihood estimator of μ and

$$(4.4) \qquad f(y|\mu_0, \sigma^2) = \sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}\|Q_\omega y\|^2\right]$$

where $Q_\omega = I - P_\omega$. Maximizing (4.4) over σ^2 yields $\hat{\sigma}_0^2 = n^{-1}\|Q_\omega y\|^2$. A similar analysis under H_1 shows that the maximum likelihood estimator of μ is $\hat{\mu}_1 = P_M y$ and $\hat{\sigma}_1^2 = n^{-1}\|Q_M y\|^2$ is the maximum likelihood estimator of σ^2. Thus the likelihood ratio test rejects for small values of the ratio

$$\Lambda(y) = \frac{f(y|\hat{\mu}_0, \hat{\sigma}_0^2)}{f(y|\hat{\mu}_1, \hat{\sigma}_1^2)} = \frac{\hat{\sigma}_0^{-n}}{\hat{\sigma}_1^{-n}} = \left(\frac{\|Q_M y\|^2}{\|Q_\omega y\|^2}\right)^{n/2}.$$

But $Q_\omega = Q_M + P_{M-\omega}$ and $Q_M P_{M-\omega} = 0$, so $\|Q_\omega y\|^2 = \|Q_M y\|^2 + \|P_{M-\omega} Y\|^2$. But rejecting for small values of $\Lambda(y)$ is equivalent to rejecting for large values of $(\Lambda(y))^{-2/n} - 1 = \|P_{M-\omega} y\|^2 / \|Q_M y\|^2$. Under H_0, $\mu \in \omega$ so $\mathcal{L}(P_{M-\omega} Y) = N(0, \sigma^2 P_{M-\omega})$ and $\mathcal{L}(Q_M Y) = N(0, \sigma^2 Q_M)$. Since $Q_M P_{M-\omega} = 0$, $Q_M Y$ and $P_{M-\omega} Y$ are independent and $\mathcal{L}(\|P_{M-\omega} Y\|) = \sigma^2 \chi_r^2$ where $r = \dim M - \dim \omega$. Also, $\mathcal{L}(\|Q_M Y\|^2) = \sigma^2 \chi_{n-k}^2$ where $k = \dim M$.

6. We use the notation of Problems 4 and 5. In the parameterization described in (iii) of Problem 4, $\beta_1 = \beta_2 = \cdots = \beta_I$ iff $\mu \in M_2$. Thus $\omega = M_2$ so $M - \omega = M_1 - M_0$. Since $M^{\perp}$ is the range of B_3 (Problem 1.15), $\|B_3 y\|^2 = \|Q_M y\|^2$, and it is clear that $\|B_3 y\|^2 = \Sigma\Sigma(y_{ij} - \bar{y}_i - \bar{y}_{.j} + \bar{y}_{..})^2$. Also, since $M - \omega = M_1 - M_0$, $P_{M-\omega} = P_{M_1} - P_{M_0}$ and $\|P_{M-\omega} y\|^2 = \|P_{M_1} y\|^2 - \|P_{M_0} y\|^2 = \Sigma_i \Sigma_j \bar{y}_{i\cdot}^2 - \Sigma_i \Sigma_j \bar{y}_{..}^2 = J\Sigma_i(\bar{y}_{i\cdot} - \bar{y}_{..})^2$.
7. Since $\mathcal{R}(X') = \mathcal{R}(X'X)$ and $X'y$ is in the range of X', there exists a $b \in R^k$ such that $X'Xb = X'y$. Now, suppose that b is any solution. First note that $P_M X = X$ since each column of X is in M. Since $X'Xb = X'y$, we have $X'[Xb - P_M y] = X'Xb - X'P_M y = X'Xb - (P_M X)'y = X'Xb - X'y = 0$. Thus the vector $v = Xb - P_M y$ is perpendicular to each column of X ($X'v = 0$) so $v \in M^{\perp}$. But $Xb \in M$, and obviously, $P_M y \in M$, so $v \in M$. Hence $v = 0$, so $Xb = P_M y$.
8. Since $I \in \gamma$, Gauss–Markov and least-squares agree iff

$$(4.5) \qquad (\alpha P_e + \beta Q_e)M \subseteq M, \qquad \text{for all } \alpha, \beta > 0.$$

But (4.5) is equivalent to the two conditions $P_e M \subseteq M$ and $Q_e M \subseteq M$.

But if $e \in M$, then $M = \text{span}\{e\} \oplus M_1$ where $M_1 \subseteq (\text{span}\{e\})^\perp$. Thus $P_e M = \text{span}\{e\} \subseteq M$ and $Q_e M = M_1 \subseteq M$, so Gauss–Markov equals least-squares. If $e \in M^\perp$, then $M \subseteq \{\text{span}\, e\}^\perp$, so $P_e M = \{0\}$ and $Q_e M = M$, so again Gauss–Markov equals least-squares. For (ii), if $e \notin M^\perp$ and $e \notin M$, then one of the two conditions $P_e M \subseteq M$ or $Q_e M \subseteq M$ is violated, so least-squares and Gauss–Markov cannot agree for all α and β. For (ii), since $M \subseteq (\text{span}\{e\})^\perp$ and $M \neq (\text{span}\{e\})^\perp$, we can write $R^n = \text{span}\{e\} \oplus M \oplus M_1$ where $M_1 = (\text{span}\{e\})^\perp - M$ and $M_1 \neq \{0\}$. Let P_1 be the orthogonal projection onto M_1. Then the exponent in the density for Y is (ignoring the factor $-\frac{1}{2}$) $(y-\mu)'(\alpha^{-1}P_e + \beta^{-1}Q_e)(y-\mu) = (P_e y + P_1 y + P_M(y-\mu))'(\alpha^{-1}P_e + \beta^{-1}Q_e)(P_e y + P_1 y + P_M(y-\mu)) = \alpha^{-1}y'P_e y + \beta^{-1}y'P_1 y + \beta^{-1}(y-\mu)'P_M(y-\mu)$ where we have used the fact that $Q_e = P_1 + P_M$ and $P_1 P_M = 0$. Since $\det(\alpha P_e + \beta Q_e) = \alpha\beta^{n-1}$, the usual arguments yields $\hat{\mu} = P_M y$, $\hat{\alpha} = y'P_e y$, and $\hat{\beta} = (n-1)^{-1}y'P_1 y$ as maximum likelihood estimators. When $M = \text{span}\{e\}$, then the maximum likelihood estimators for (α, μ) do not exist—other than the solution $\hat{\mu} = P_e y$ and $\hat{\alpha} = 0$ (which is outside the parameter space). The whole point is that when $e \in M$, you must have replications to estimate α when the covariance structure is $\alpha P_e + \beta Q_e$.

9. Define the inner product $(\cdot, \cdot)$ on R^n by $(x, y) = x'\Sigma_1^{-1}y$. In the inner product space $(R^n, (\cdot, \cdot))$, $\mathcal{E}Y = X\beta$ and $\text{Cov}(Y) = \sigma^2 I$. The transformation P defined by the matrix $X(X'\Sigma_1^{-1}X)^{-1}X'\Sigma_1^{-1}$ satisfies $P^2 = P$ and is self-adjoint in $(R^n, (\cdot, \cdot))$. Thus P is an orthogonal projection onto its range, which is easily shown to be the column space of X. The Gauss–Markov Theorem implies that $\hat{\mu} = PY$ as claimed. Since $\mu = X\beta$, $X'\mu = X'X\beta$ so $\beta = (X'X)^{-1}X'\mu$. Hence $\hat{\beta} = (X'X)^{-1}X'\hat{\mu}$, which is just the expression given.

10. For (i), each $\Gamma \in \mathcal{O}(V)$ is nonsingular so $\Gamma(M) \subseteq M$ is equivalent to $\Gamma(M) = M$—hence $\Gamma^{-1}(M) = M$ and $\Gamma^{-1} = \Gamma'$. Parts (ii) and (iii) are easy. To verify (iv), $t_0(c\Gamma Y + x_0) = P_M(c\Gamma Y + x_0) = cP_M\Gamma Y + x_0 = c\Gamma P_M Y + x_0 = c\Gamma t_0(Y) + x_0$. The identity $P_M\Gamma = \Gamma P_M$ for $\Gamma \in \mathcal{O}_M(V)$ was used to obtain the third equality. For (v), first set $\Gamma = I$ and $x_0 = -P_M y$ to obtain

$$t(y) = t(Q_M y) + P_M y. \tag{4.6}$$

Then to calculate t, we need only know t for vectors $u \in M^\perp$ as $Q_M y \in M^\perp$. Fix $u \in M^\perp$ and let $z = t(u)$ so $z \in M$ by assumption. Then there exists a $\Gamma \in \mathcal{O}_M(V)$ such that $\Gamma u = u$ and $\Gamma z = -z$. For this Γ, we have $z = t(u) = t(\Gamma u) = \Gamma t(u) = \Gamma z = -z$ so $z = 0$. Hence $t(u) = 0$ for all $u \in M^\perp$ and the result follows.

11. Part (i) follows by showing directly that the regression subspace M is invariant under each $I_n \otimes A$. For (ii), an element of M has the form $\mu = \{Z_1\beta_1, Z_2\beta_2\} \in \mathcal{L}_{2,n}$ for some $\beta_1 \in R^k$ and $\beta_2 \in R^k$. To obtain an example where M is not invariant under all $I_n \otimes \Sigma$, take $k = 1$, $Z_1 = \varepsilon_1$, and $Z_2 = \varepsilon_2$ so μ is

$$\mu = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

That the set of such μ's is not invariant under all $I_n \otimes \Sigma$ is easily verified. When $Z_1 = Z_2$, then $\mu = Z_1 B$ where B is $k \times 2$ with ith column β_i, $i = 1, 2$. Thus Example 4.4 applies. For (iii), first observe that Z_1 and Z_2 have the same column space (when they are of full rank) iff $Z_2 = Z_1 C$ where C is $k \times k$ and nonsingular. Now, apply part (ii) with β_2 replaced by $C\beta_2$, so M is the set of μ's of the form $\mu = Z_1 B$ where $B \in \mathcal{L}_{2,k}$.

CHAPTER 5

1. Let $a_1, \ldots, a_p$ be the columns of A and apply Gram–Schmidt to these vectors in the order $a_p, a_{p-1}, \ldots, a_1$. Now argue as in Proposition 5.2.
2. Follows easily from the uniqueness of $F(S)$.
3. Just modify the proof of Proposition 5.4.
4. Apply Proposition 5.7
5. That F is one-to-one and onto follows from Proposition 5.2. Given $A \in \mathcal{L}^0_{p,n}$, $F^{-1}(A) \in \mathcal{F}_{p,n} \times G_u^+$ is the pair (ψ, U) where $A = \psi U$. For (ii), $F(\Gamma\psi, UT') = \Gamma\psi UT' = (\Gamma \otimes T)(\psi U) = (\Gamma \otimes T)(F(\psi, U))$. If $F^{-1}(A) = (\psi, U)$, then $A = \psi U$ and ψ and U are unique. Then $(\Gamma \otimes T)A = \Gamma A T' = \Gamma\psi UT'$ and $\Gamma\psi \in \mathcal{F}_{p,n}$ and $UT' \in G_U^+$. Uniqueness implies that $F^{-1}(\Gamma\psi UT') = (\Gamma\psi, UT')$.
6. When $D_g(x_0)$ exists, it is the unique $n \times n$ matrix that satisfies

$$\lim_{x \to x_0} \frac{\|g(x) - g(x_0) - D_g(x_0)(x - x_0)\|}{\|x - x_0\|} = 0. \tag{5.3}$$

But by assumption, (5.3) is satisfied by A (for $D_g(x_0)$). By definition $J_g(x_0) = \det(D_g(x_0))$.

7. With t_{ii} denoting the ith diagonal element of T, the set $\{T|t_{ii} > 0\}$ is open since the function $T \to t_{ii}$ is continuous on V to R^1. But $G_T^+ = \cap_1^p\{T|t_{ii} > 0\}$, which is open. That g has the given representation is just a matter of doing a little algebra. To establish the fact that $\lim_{x\to 0}(\|R(x)\|/\|x\|) = 0$, we are free to use any norm we want on V and $\mathbb{S}_p^+$ (all norms defined by inner products define the same topology). Using the trace inner product on V and $\mathbb{S}_p^+$, $\|R(x)\|^2 = \|xx'\|^2 = \operatorname{tr} xx'xx'$ and $\|x\|^2 = \operatorname{tr} xx'$, $x \in V$. But for $S \geqslant 0$, $\operatorname{tr} S^2 \leqslant (\operatorname{tr} S)^2$ so $\|R(x)\|/\|x\| \leqslant \operatorname{tr} xx'$, which converges to zero as $x \to 0$. For (iii), write $S = L(x)$, string the S coordinates out as a column vector in the order $s_{11}, s_{21}, s_{22}, s_{31}, s_{32}, s_{33}, \ldots,$ and string the x coordinates out in the same order. Then the matrix of L is lower triangular and its determinant is easily computed by induction. Part (iv) is immediate from Problem 6.
8. Just write out the equations $SS^{-1} = I$ in terms of the blocks and solve.
9. That $P^2 = P$ is easily checked. Also, some algebra and Problem 8 show that $(Pu, v) = (u, Pv)$ so P is self-adjoint in the inner product $(\cdot,\cdot)$. Thus P is an orthogonal projection on $(R^p, (\cdot,\cdot))$. Obviously,

$$R(P) = \left\{x|x = \begin{pmatrix} y \\ z \end{pmatrix}, z = 0\right\}.$$

Since

$$Px = \begin{pmatrix} y - \Sigma_{12}\Sigma_{22}^{-1}z \\ 0 \end{pmatrix},$$

$$\|Px\|^2 = (Px, Px) = \begin{pmatrix} y - \Sigma_{12}\Sigma_{22}^{-1}z \\ 0 \end{pmatrix}'\Sigma^{-1}\begin{pmatrix} y - \Sigma_{12}\Sigma_{22}^{-1}z \\ 0 \end{pmatrix}$$

$$= \left(y - \Sigma_{12}\Sigma_{22}^{-1}z\right)'\Sigma^{11}\left(y - \Sigma_{12}\Sigma_{22}^{-1}z\right).$$

A similar calculation yields $\|(I - P)x\|^2 = z'\Sigma_{22}^{-1}z$. For (iii), the exponent in the density of X is $-\frac{1}{2}(x, x) = -\frac{1}{2}\|Px\|^2 - \frac{1}{2}\|(I - P)x\|^2$. Marginally, Z is $N(0, \Sigma_{22})$, so the exponent in Z's density is $-\frac{1}{2}\|(I - P)x\|^2$. Thus dividing shows that the exponent in the conditional density of Y given Z is $-\frac{1}{2}\|Px\|^2$, which corresponds to a normal distribution with mean $\Sigma_{12}\Sigma_{22}^{-1}Z$ and covariance $(\Sigma^{11})^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.
10. On G_T^+, for $j < i$, t_{ij} ranges from $-\infty$ to $+\infty$ and each integral contributes $\sqrt{2\pi}$—there are $p(p-1)/2$ of these. For $j = i$, t_{ii} ranges

from 0 to ∞ and the change of variable $u_{ii} = t_{ii}^2/2$ shows that the integral over t_{ii} is $(\sqrt{2})^{r-i-1}\Gamma((r-i+1)/2)$. Hence the integral is equal to

$$\pi^{(p(p-1))/4}2^{(p(p-1))/4}2^{1/2\Sigma(r-i-1)}\prod_1^p \Gamma\left(\frac{r-i+1}{2}\right),$$

which is just $2^{-p}c(r, p)$.

CHAPTER 6

1. Each $g \in Gl(V)$ maps a linearly independent set into a linearly independent set. Thus $g(M) \subseteq M$ implies $g(M) = M$ as $g(M)$ and M have the same dimension. That $G(M)$ is a group is clear. For (ii),

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} y \\ 0 \end{pmatrix} \in M \qquad \text{for } y \in R^q$$

iff $g_{21}y = 0$ for $y \in R^q$ iff $g_{21} = 0$. But

$$\begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix}$$

is nonsingular iff both g_{11} and g_{22} are nonsingular. That G_1 and G_2 are subgroups of $G(M)$ is obvious. To show G_2 is normal, consider $h \in G_2$ and $g \in G(M)$. Then

$$ghg^{-1} = \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix}\begin{pmatrix} h_{11} & h_{12} \\ 0 & I_r \end{pmatrix}\begin{pmatrix} g_{11}^{-1} & -g_{11}^{-1}g_{12}g_{22}^{-1} \\ 0 & g_{22}^{-1} \end{pmatrix}$$

has its 2, 2 element I_r, so is in G_2. For (iv), that $G_1 \cap G_2 = \{I\}$ is clear. Each $g \in G$ can be written as

$$g = \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ 0 & g_{22} \end{pmatrix}\begin{pmatrix} g_{11} & g_{12} \\ 0 & I_r \end{pmatrix},$$

which has the form $g = hk$ with $h \in G_1$ and $k \in G_2$. The representation is unique as $G_1 \cap G_2 = \{I\}$. Also, $g_1g_2 = h_1k_1h_2k_2 = h_1h_2h_2^{-1}k_1h_2k_2 = h_3k_3$ by the uniqueness of the representation.

2. $G(M)$ does not act transitively on $V - \{0\}$ since the vector $\binom{y}{0}$, $y \neq 0$ remains in M under the action of each $g \in G$. To show $G(M)$ is

transitive on $V \cap M^c$, consider

$$x_i = \begin{pmatrix} y_i \\ z_i \end{pmatrix}, \qquad i = 1, 2$$

with $z_1 \neq 0$ and $z_2 \neq 0$. It is easy to argue there is a $g \in G(M)$ such that $gx_1 = x_2$ (since $z_1 \neq 0$ and $z_2 \neq 0$).

3. Each $n \times n$ matrix $\Gamma \in \mathcal{O}_n$ can be regarded as an n^2-dimensional vector. A sequence $\{\Gamma_j\}$ converges to a point $x \in R^m$ iff each element of Γ_j converges to the corresponding element of x. It is clear that the limit of a sequence of orthogonal matrices is another orthogonal matrix. To show $\mathcal{O}_n$ is a topological group, it must be shown that the map $(\Gamma, \psi) \to \Gamma\psi'$ is continuous from $\mathcal{O}_n \times \mathcal{O}_n$ to $\mathcal{O}_n$—this is routine. To show $\chi(\Gamma) = 1$ for all Γ, first observe that $H = \{\chi(\Gamma) | \Gamma \in \mathcal{O}_n\}$ is a subgroup of the multiplicative group $(0, \infty)$ and H is compact as it is the continuous image of a compact set. Suppose $r \in H$ and $r \neq 1$. Then $r^j \in H$ for $j = 1, 2, \ldots$ as H is a group, but $\{r^j\}$ has no convergent subsequence—this contradicts the compactness of H. Hence $r = 1$.
4. Set $x = e^u$ and $\xi(u) = \log \chi(e^u)$, $u \in R^1$. Then $\xi(u_1 + u_2) = \xi(u_1) + \xi(u_2)$ so ξ is a continuous homomorphism on R^1 to R^1. It must be shown that $\xi(u) = \nu u$ for some fixed real ν. This follows from the solution to Problem 6 below in the special case that $V = R^1$.
5. This problem is easy, but the result is worth noting.
6. Part (i) is easy and for part (ii), all that needs to be shown is that ϕ is linear. First observe that

$$\phi(v_1 + v_2) = \phi(v_1) + \phi(v_2) \tag{6.6}$$

so it remains to verify that $\phi(\lambda v) = \lambda\phi(v)$ for $\lambda \in R^1$. (6.6) implies $\phi(0) = 0$ and $\phi(nv) = n\phi(v)$ for $n = 1, 2, \ldots$. Also, $\phi(-v) = -\phi(v)$ follows from (6.6). Setting $w = nv$ and dividing by n, we have $\phi(w/n) = (1/n)\phi(w)$ for $n = 1, 2, \ldots$. Now $\phi((m/n)v) = m\phi((1/n)v) = (m/n)\phi(v)$ and by continuity, $\phi(\lambda v) = \lambda\phi(v)$ for $\lambda > 0$. The rest is easy.
7. Not hard with the outline given.
8. By the spectral theorem, every rank r orthogonal projection can be written $\Gamma x_0 \Gamma'$ for some $\Gamma \in \mathcal{O}_n$. Hence transitivity holds. The equation $\Gamma x_0 \Gamma' = x_0$ holds for $\Gamma \in \mathcal{O}_n$ iff Γ has the form

$$\Gamma = \begin{pmatrix} \Gamma_{11} & 0 \\ 0 & \Gamma_{22} \end{pmatrix} \in \mathcal{O}_n,$$

and this gives the isotropy subgroup of x_0. For $\Gamma \in \mathcal{O}_n$, $\Gamma x_0 \Gamma' = \Gamma x_0(\Gamma x_0)'$ and Γx_0 has the form $(\psi 0)$ where $\psi : n \times r$ has columns that are the first r columns of Γ. Thus $\Gamma x_0 \Gamma' = \psi\psi'$. Part (ii) follows by observing that $\psi_1\psi_1' = \psi_2\psi_2'$ if $\psi_1 = \psi_2\Delta$ for some $\Delta \in \mathcal{O}_r$.

9. The only difficulty here is (iii). The problem is to show that the only continuous homomorphisms χ on G_2 to (∞, ∞) are t_{pp}^{α} for some real α. Consider the subgroups G_3 and G_4 of G_2 given by

$$G_3 = \left\{ \begin{pmatrix} I_{p-1} & 0 \\ x & 1 \end{pmatrix} \middle| x' \in R^{p-1} \right\}, \qquad G_4 = \left\{ \begin{pmatrix} I_{p-1} & 0 \\ 0 & u \end{pmatrix} \middle| u \in (0, \infty) \right\}.$$

The group G_3 is isomorphic to R^{p-1} so the only homomorphisms are $x \to \exp[\Sigma_1^{p-1} a_i x_i]$ and G_4 is isomorphic to $(0, \infty)$ so the only homomorphisms are $u \to u^{\alpha}$ for some real α. For $k \in G_2$, write

$$k = \begin{pmatrix} I_{p-1} & 0 \\ x & u \end{pmatrix} = \begin{pmatrix} I_{p-1} & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} I_{p-1} & 0 \\ 0 & u \end{pmatrix}$$

so $\chi(k) = \exp[\Sigma a_i x_i] u^{\alpha}$. Now, use the condition $\chi(k_1 k_2) = \chi(k_1) \cdot \chi(k_2)$ to conclude $a_1 = a_2 = \cdots = a_{p-1} = 0$ so χ has the claimed form.

10. Use (6.4) to conclude that

$$I_{\gamma} = 2^p (\sqrt{2\pi})^{np} \omega(n, p) \int_{G_U^+} \prod_1^p U_{ii}^{2\gamma + n - i} \exp\left[-\tfrac{1}{2} \sum\sum_{i \leqslant j} U_{ij}^2 \right] dU$$

and then use Problem 5.10 to evaluate the integral over G_U^+. You will find that, for $2\gamma + n > p - 1$, the integral is finite and is $I_{\gamma} = (\sqrt{2\pi})^{np} \omega(n, p)/\omega(2\gamma + n, p)$. If $2\gamma + n \leqslant p - 1$, the integral diverges.

11. Examples 6.14 and 6.17 give Δ_r for $G(M)$ and all the continuous homomorphisms for $G(M)$. Pick $x_0 \in R^p \cap M^c$ to be

$$x_0 = \begin{pmatrix} 0 \\ z_0 \end{pmatrix}$$

where $z_0' = (1, 0, \ldots, 0)$, $z_0 \in R^r$. Then H_0 consists of those g's with the first column of g_{12} being 0 and the first column of g_{22} being z_0. To apply Theorem 6.3, all that remains is to calculate the right-hand modulus of H_0—say Δ_r^0. This is routine given the calculations of Examples 6.14 and 6.17. You will find that the only possible multipliers are $\chi(g) = |g_{11}||g_{33}|$ and Lebesgue measure on $R^p \cap M^c$ is the only (up to a positive constant) invariant measure.

12. Parts (i), (ii), (iii), and (iv) are routine. For (v), $J_1(f) = \int f(x)\mu(dx)$ and $J_2(f) = \int f(\tau^{-1}(y))\nu(dy)$ are both invariant integrals on $\mathcal{K}(\mathfrak{X})$. By Theorem 6.3, $J_1 = kJ_2$ for some constant k. To find k, take $f(x) = (\sqrt{2\pi})^{-n}s^n(x)\exp[-\frac{1}{2}x'x]$ so $J_1(f) = 1$. Since $s(\tau^{-1}(y)) = v$ for $y = (u, v, w)$,

$$J_2(f) = (\sqrt{2\pi})^{-n}\int_y v^n \exp\left[-\tfrac{1}{2}v^2 - \tfrac{1}{2}nu^2\right] du\frac{dv}{v^2}\nu(dw)$$

$$= \frac{1}{2}\frac{\Gamma((n-1)/2)}{(\sqrt{\pi})^{n-1}} = \frac{1}{k}.$$

For (vi), the expected value of any function of $\bar{x}$ and $s(x)$, say $q(\bar{x}, s(x))$ is

$$\mathcal{E}q(\bar{x}, s(x)) = \int q(\bar{x}, s(x))f(x)s^n(x)\mu(dx)$$

$$= k\int q(u, v)f(\tau^{-1}(u, v, w))v^n\, du\frac{dv}{v^2}\nu(dw)$$

$$= k\int q(u, v)\frac{v^{n-2}}{\sigma^2}h\left(\frac{v^2}{\sigma^2} + \frac{n(u-\delta)^2}{\sigma^2}\right) du\, dv.$$

Thus the joint density of $\bar{x}$ and $s(x)$ is

$$p(u, v) = \frac{kv^{n-2}}{\sigma^n}h\left(\frac{v^2}{\sigma^2} + \frac{n(u-\delta)^2}{\sigma^2}\right) \qquad \text{(with respect to } du\, dv).$$

13. We need to show that, with $Y(X) = X/\|X\|$, $P\{\|X\| \in B, Y \in C\} = P\{\|X\| \in B\}P\{Y \in C\}$. If $P\{\|X\| \in B\} = 0$, the above is obvious. If not, set $\nu(C) = P\{Y \in C, \|X\| \in B\}/P\{\|X\| \in B\}$ so ν is a probability measure on the Borel sets of $\{y | \|y\| = 1\} \subseteq R^n$. But the relation $\phi(\Gamma x) = \Gamma\phi(x)$ and the $\mathcal{O}_n$ invariance of $\mathcal{L}(X)$ implies that ν is an $\mathcal{O}_n$-invariant probability measure and hence is unique —(for all Borel B)—namely, ν is uniform probability measure on $\{y | \|y\| = 1\}$.

14. Each $x \in \mathfrak{X}$ can be uniquely written as gy with $g \in \mathcal{P}_n$ and $y \in \mathcal{Y}$ (of course, y is the order statistic of x). Define $\mathcal{P}_n$ acting on $\mathcal{P}_n \times \mathcal{Y}$ by

$g(P, y) = (gP, y)$. Then $\phi^{-1}(gx) = g\phi^{-1}(x)$. Since $P(gx) = gP(x)$, the argument used in Problem 13 shows that $P(X)$ and $Y(X)$ are independent and $P(X)$ is uniform on $\mathcal{P}_n$.

CHAPTER 7

1. Apply Propositions 7.5 and 7.6.
2. Write $X = \psi U$ as in Proposition 7.3 so ψ and U are independent. Then $P(X) = \psi\psi'$ and $S(X) = U'U$ and the independence is obvious.
3. First, write Q in the form

$$Q = M'\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}M$$

where M is $n \times n$ and nonsingular. Since M is nonsingular, it suffices to show that $(M^{-1}(A))^c$ has measure zero. Write $x = \binom{\dot{x}}{\ddot{x}}$ where $\dot{x}$ is $r \times p$. It then suffices to show that $B^c = \{x | x \in \mathcal{L}_{p,n}, \text{rank}(\dot{x}) = p\}^c$ has measure zero. For this, use the argument given in Proposition 7.1.
4. That the ϕ's are the only equivariant functions follows as in Example 7.6.
5. Part (i) is obvious. For (ii), just observe that knowledge of F_n allows you to write down the order statistic and conversely.
6. Parts (i) and (ii) are clear. For (iii), write $x = Px + Qx$. If t is equivariant $t(x + y) = t(x) + y$, $y \in M$. This implies that $t(Qx) = t(x) + Px$ (pick $y = Px$). Thus $t(x) = Px + t(Qx)$. Since $Q = I - P$, $Qx \in M^{\perp}$, so $BQx = Qx$ for any B with $(B, y) \in G$. Since $t(Qx) \in M$, pick B such that $Bx = -x$ for $x \in M$. The equivariance of t then gives $t(Qx) = t(BQx) = Bt(Qx) = -t(Qx)$, so $t(Qx) = 0$.
7. Part (i) is routine as is the first part of (ii) (use Problem 6). An equivariant estimator of σ^2 must satisfy $t(a\Gamma x + b) = a^2 t(x)$. G acts transitively on $\mathcal{X}$ and $\overline{G}$ acts transitively on $(0, \infty)$ ($\mathcal{Y}$ for this case) so Proposition 7.8 and the argument given in Example 7.6 apply.
8. When $X \in \mathcal{X}$ with density $f(x'x)$, then $Y = X\Sigma^{1/2} = (I_n \otimes \Sigma^{1/2})X$ has density $f(\Sigma^{-1/2}x'x\Sigma^{-1/2})$ since $dx/|x'x|^{n/2}$ is invariant under $x \to xA$ for $A \in Gl_p$. Also, when X has density f, then $\mathcal{L}((\Gamma \otimes \Delta)X) = \mathcal{L}(X)$ for all $\Gamma \in \mathcal{O}_n$ and $\Delta \in \mathcal{O}_p$. This implies (see Proposition 2.19) that $\text{Cov}(X) = cI_n \otimes I_p$ for some $c > 0$. Hence $\text{Cov}((I_n \otimes \Sigma^{1/2})X) = cI_n \otimes \Sigma$. Part (ii) is clear and (iii) follows from Proposition 7.8 and Example 7.6. For (iv), the definition of C_0 and the assumption on f

imply $f(\Gamma C_0\Gamma') = f(C_0\Gamma'\Gamma) = f(C_0)$ for each $\Gamma \in \mathcal{O}_p$. The uniqueness of C_0 implies $C_0 = \alpha I_p$ for some $\alpha > 0$. Thus the maximum likelihood estimator of Σ must be $\alpha X'X$ (see Proposition 7.12 and Example 7.10).

9. If $\mathcal{L}(X) = P_0$, then $\mathcal{L}(\|X\|)$ is the same whenever $\mathcal{L}(X) \in \{P|P = gP_0, g \in \mathcal{O}(V)\}$ since $x \to \|x\|$ is a maximal invariant under the action of $\mathcal{O}(V)$ on V. For (ii), $\mathcal{L}(\|X\|)$ depends on μ through $\|\mu\|$.
10. Write $V = \omega \oplus (M - \omega) \oplus M^{\perp}$. Remove a set of Lebesgue measure zero from V and show the F ratio is a maximal invariant under the group action $x \to a\Gamma x + b$ where $a > 0$, $b \in \omega$, and $\Gamma \in \mathcal{O}(V)$ satisfies $\Gamma(\omega) \subseteq \omega$, $\Gamma(M - \omega) \subseteq (M - \omega)$. The group action on the parameter (μ, σ^2) is $\mu \to a\Gamma\mu + b$ and $\sigma^2 \to a^2\sigma^2$. A maximal invariant parameter is $\|P_{M-\omega}\mu\|^2/\sigma^2$, which is zero when $\mu \in \omega$.
11. The statistic V is invariant under $x_i \to Ax_i + b$, $i = 1,\ldots, n$, where $b \in R^p$, $A \in Gl_p$, and $\det A = 1$. The model is invariant under this group action where the induced group action on (μ, Σ) is $\mu \to A\mu + b$ and $\Sigma \to A\Sigma A'$. A direct calculation shows $\theta = \det(\Sigma)$ is a maximal invariant under the group action. Hence the distribution of V depends on (μ, Σ) only through θ.
12. For (i), if $h \in G$ and $B \in \mathcal{B}$, $(hP)(B) = P(h^{-1}B) = \int_G(g\overline{Q})(h^{-1}B)\mu(dg) = \int_G\overline{Q}(g^{-1}h^{-1}B)\mu(dg) = \int_G\overline{Q}((hg)^{-1}B)\mu(dg) = \int\overline{Q}(g^{-1}B)\mu(dg) = P(B)$, so $hP = P$ for $h \in G$ and P is G invariant. For (ii), let Q be the distribution described in Proposition 7.16 (ii), so if $\mathcal{L}(X) = P$, then $\mathcal{L}(X) = \mathcal{L}(UY)$ where U is uniform on G and is independent of Y. Thus for any bounded $\mathcal{B}$-measurable function f,

$$\int f(x)P(dx) = \int_G\int_{\mathcal{Y}} f(gy)\mu(dg)Q(dy) = \int_G\int_{\mathcal{X}} f(gx)\mu(dg)\overline{Q}(dx).$$

Set $f = I_B$ and we have $P(B) = \int_G\overline{Q}(g^{-1}B)\mu(dg)$ so (7.1) holds.

13. For $y \in \mathcal{Y}$ and $B \in \mathcal{B}$, define $R(B|y)$ by $R(B|y) = \int_G I_B(gy)\mu(dg)$. For each y, $R(\cdot|y)$ is a probability measure on $(\mathcal{X}, \mathcal{B})$ and for fixed B, $R(B|\cdot)$ is $(\mathcal{Y}, \mathcal{C})$ measurable. For $P \in \mathcal{P}$, (ii) of Proposition 7.16 shows that

$$\int h(x)P(dx) = \int_{\mathcal{Y}}\int_G h(gy)\mu(dg)Q(dy). \tag{7.2}$$

But by definition of $R(\cdot|\cdot)$, $\int_G h(gy)\mu(dg) = \int_{\mathcal{X}} h(x)R(dx|y)$, so (7.2) becomes

$$\int_{\mathcal{X}} h(x)P(dx) = \int_{\mathcal{Y}}\int_{\mathcal{X}} h(x)R(dx|y)Q(dy).$$

This shows that $R(\cdot|y)$ serves as a version of the conditional distribution of X given $\tau(X)$. Since R does not depend on $P \in \mathcal{P}$, $\tau(X)$ is sufficient.

14. For (i), that $t(gx) = g \circ t(x)$ is clear. Also, $X - \bar{X}e = Q_e X$, which is $N(0, Q_e)$ so is ancillary. For (ii), $\mathcal{E}(f(X_1)|\bar{X} = t) = \mathcal{E}(f(X_1 - \bar{X} + \bar{X})|\bar{X} = t) = \mathcal{E}(f(\varepsilon_1' Z(X) + \bar{X})|\bar{X} = t)$ since $Z(X)$ has coordinates $X_i - \bar{X}$, $i = 1, \ldots, n$. Since Z and $\bar{X}$ are independent, this last conditional expectation (given $\bar{X} = t$) is just the integral over the distribution of Z with $\bar{X} = t$. But $\varepsilon_1' Z(X) = X_1 - \bar{X}$ is $N(0, \delta^2)$ so the claimed integral expression holds. When $f(x) = 1$ for $x \leqslant u_0$ and 0 otherwise, the integral is just $\Phi((u_0 - t)/\delta)$ where Φ is the normal cumulative distribution function.

15. Let B be the set $(-\infty, u_0]$ so $I_B(X_1)$ is an unbiased estimator of $h(a, b)$ when $\mathcal{L}(X) = (a, b)P_0$. Thus $\hat{h}(t(X)) = \mathcal{E}(I_B(X_1)|t(X))$ is an unbiased estimator of $h(a, b)$ based on $t(X)$. To compute $\hat{h}$, we have $\mathcal{E}(I_B(X_1)|t(X)) = P\{X_1 \leqslant u_0|t(X)\} = P\{(X_1 - \bar{X})/s \leqslant (u_0 - \bar{X})/s|(s, \bar{X})\}$. But $(X_1 - \bar{X})/s \equiv Z_1$ is the first coordinate of $Z(X)$ so is independent of $(s, \bar{X})$. Thus $\hat{h}(s, \bar{X}) = P_{Z_1}\{Z_1 \leqslant (u_0 - \bar{X})/s\} = F((u_0 - \bar{X})/s)$ where F is the distribution function of the first coordinate of Z. To find F, first observe that Z takes values in $\mathcal{Z} = \{x|x \in R^n, x'e = 0, \|x\| = 1\}$ and the compact group $\mathcal{O}_n(e)$ acts transitively on $\mathcal{Z}$. Since $Z(\Gamma X) = \Gamma Z(X)$ for $\Gamma \in \mathcal{O}_n(e)$, it follows that Z has a uniform distribution on $\mathcal{Z}$ (see the argument in Example 7.19). Let U be $N(0, I_n)$ so Z has the same distribution as $Q_e U/\|Q_e U\|$ and $\mathcal{L}(Z_1) = \mathcal{L}(\varepsilon_1' Q_e U/\|Q_e U\|^2) = \mathcal{L}((Q_e\varepsilon_1)' Q_e U/\|Q_e U\|^2)$. Since $\|Q_e \varepsilon_1\|^2 = (n - 1)/n$ and $Q_e U$ is $N(0, Q_e)$, it follows that $\mathcal{L}(Z_1) = \mathcal{L}(((n - 1)/n)^{1/2} W_1)$ where $W_1 = U_1/(\Sigma_1^{n-1} U_i^2)^{1/2}$. The rest is a routine computation.

16. Part (i) is obvious and (ii) follows from

$$\begin{aligned}(7.3)\quad \mathcal{E}(f(X)|\tau(X) = g) &= \mathcal{E}\big(f\big(\tau(X)(\tau(X))^{-1}X\big)|\tau(X) = g\big)\\ &= \mathcal{E}(f(\tau(X)Z(X))|\tau(X) = g).\end{aligned}$$

Since $Z(X)$ and $\tau(X)$ are independent and $\tau(X) = g$, the last member of (7.3) is just the expectation over Z of $f(gZ)$. Part (iii) is just an application and Q_0 is the uniform distribution on $\mathcal{F}_{p,n}$. For (iv), let B be a fixed Borel set in R^p and consider the parametric function

$h(\Sigma) = P_\Sigma(X_1 \in B) = \int I_B(x)(\sqrt{2\pi})^{-p}|\Sigma|^{-1/2}\exp[-\frac{1}{2}x'\Sigma^{-1}x]dx$, where X_1' is the first row of X. Since $\tau(X)$ is a complete sufficient statistic, the MVUE of $h(\Sigma)$ is

(7.4)

$$\hat{h}(T) = \mathcal{E}(I_B(X_1)|\tau(X) = T) = P\{T(\tau(X))^{-1}X_1 \in B|\tau(X) = T\}.$$

But $Z_1' = (\tau^{-1}(X)X_1)'$ is the first row of $Z(X)$ so is independent of $\tau(X)$. Hence $\hat{h}(T) = P_1\{Z_1 \in T^{-1}(B)\}$ where P_1 is the distribution of Z_1 when Z has a uniform distribution on $\mathcal{F}_{p,n}$. Since Z_1 is the first p coordinates of a random vector that is uniform on $\{x|\|x\| = 1, x \in R^n\}$, it follows that Z_1 has a density $\psi(\|u\|^2)$ for $u \in R^p$ where ψ is given by

$$\psi(v) = \begin{cases} c(1-v)^{(n-p-2)/2} & 0 < v < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $c = \Gamma(n/2)/\pi^{p/2}\Gamma((n-p)/2)$. Therefore $\hat{h}(T) = \int_{R^p} I_B(Tu)\psi(\|u\|^2)du = (\det T)^{-1}\int_{R^p} I_B(u)\psi(\|T^{-1}u\|^2)du$. Now, let B shrink to the point u_0 to get that $(\det T)^{-1}\psi(\|T^{-1}u_0\|^2)$ is the MVUE for $(\sqrt{2\pi})^{-p}|\Sigma|^{-1/2}\exp[-\frac{1}{2}u_0'\Sigma^{-1}u_0]$.

CHAPTER 8

1. Make a change of variables to r, $x_1 = s_{11}/\sigma_{11}$ and $x_2 = s_{22}/\sigma_{22}$, and then integrate out x_1 and x_2. That $p(r|\rho)$ has the claimed form follows by inspection. Karlin's Lemma (see Appendix) implies that $\psi(\rho r)$ has a monotone likelihood ratio.
3. For $\alpha = 1/2, \ldots, (p-1)/2$, let $X_1, \ldots, X_r$ be i.i.d. $N(0, I_p)$ with $r = 2\alpha$. Then $S = X_i X_i'$ has ϕ_α as its characteristic function. For $\alpha > (p-1)/2$, the function $p_\alpha(s) = k(\alpha)|s|^\alpha \exp[-\frac{1}{2}\operatorname{tr} s]$ is a density with respect to $ds/|s|^{(p+1)/2}$ on $\mathcal{S}_p^+$. The characteristic function of p_α is ϕ_α. To show that $\phi_\alpha(\Sigma A)$ is a characteristic function, let S satisfy $\mathcal{E}\exp(i\langle A, S\rangle) = \phi_\alpha(A) = |I_p - 2iA|^\alpha$. Then $\Sigma^{1/2}S\Sigma^{1/2}$ has $\phi_\alpha(\Sigma A)$ as its characteristic function.
4. $\mathcal{L}(S) = \mathcal{L}(\Gamma S \Gamma')$ implies that $A = \mathcal{E}S$ satisfies $A = \Gamma A \Gamma'$ for all $\Gamma \in \mathcal{O}_p$. This implies $A = cI_p$ for some constant c. Obviously, $c = \mathcal{E}s_{11}$. For (ii) $\operatorname{var}(\operatorname{tr} DS) = \operatorname{var}(\Sigma_1^p d_i s_{ii}) = \Sigma_1^p d_i^2 \operatorname{var}(s_{ii}) + \Sigma\Sigma_{i \neq j} d_i d_j \operatorname{cov}(s_{ii}, s_{jj})$. Noting that $\mathcal{L}(S) = \mathcal{L}(\Gamma S \Gamma')$ for $\Gamma \in \mathcal{O}_p$, and in particular for permutation matrices, it follows that $\gamma = \operatorname{var}(s_{ii})$ does not depend on i and $\beta = \operatorname{cov}(s_{ii}, s_{jj})$ does not depend on i and j $(i \neq j)$. Thus $\operatorname{var}\langle D, S\rangle =$

$\gamma\Sigma_1^p d_i^2 + \beta\Sigma\Sigma_{i\neq j} d_i d_j = (\gamma - \beta)\Sigma_1^p d_i^2 + \beta(\Sigma_1^p d_i)^2$. For (iii), write $A \in \mathbb{S}_p$ as $\Gamma D \Gamma'$ so $\operatorname{var}\langle A, S\rangle = \operatorname{var}\langle \Gamma D\Gamma', S\rangle = \operatorname{var}\langle D, \Gamma' S\Gamma\rangle = \operatorname{var}\langle D, S\rangle = (\gamma - \beta)\Sigma_1^p d_i^2 + \beta(\Sigma_1^p d_i)^2 = (\gamma - \beta)\operatorname{tr} A^2 + \beta(\operatorname{tr} A)^2 = (\gamma - \beta)\langle A, A\rangle + \beta\langle I, A\rangle^2$. With $T = (\gamma - \beta)I_p \otimes I_p + \beta I_p \square I_p$, it follows that $\operatorname{var}\langle A, S\rangle = \langle A, TA\rangle$, and since T is self-adjoint, this implies that $\operatorname{Cov}(S) = T$.

5. Use Proposition 7.6.
6. Immediate from Problem 3.
7. For (i), it suffices to show that $\mathcal{L}((ASA')^{-1}) = W((A\Lambda A')^{-1}, r, \nu + r - 1)$. Since $\mathcal{L}(S^{-1}) = W(\Lambda^{-1}, p, \nu + p - 1)$, Proposition 8.9 implies that desired result. (ii) follows immediately from (i). For (iii), (i) implies $\tilde{S} = \Lambda^{-1/2}S\Lambda^{-1/2}$ is $IW(I_p, p, \nu)$ and $\mathcal{L}(\tilde{S}) = \mathcal{L}(\Gamma\tilde{S}\Gamma')$ for all $\Gamma \in \mathcal{O}_p$. Now, apply Problem 4 to conclude that $\mathcal{E}\tilde{S} = cI_p$ where $c = \mathcal{E}\tilde{s}_{11}$. That $c = (\nu - 2)^{-1}$ is an easy application of (i). Hence $(\nu - 2)^{-1}I_p = \mathcal{E}\tilde{S} = \Lambda^{-1/2}(\mathcal{E}S)\Lambda^{-1/2}$ so $\mathcal{E}S = (\nu - 2)^{-1}\Lambda$. Also, $\operatorname{Cov}\tilde{S} = (\gamma - \beta)I_p \otimes I_p + \beta I_p \square I_p$ as in Problem 4. Thus $\operatorname{Cov}(\tilde{S}) = (\Lambda^{1/2} \otimes \Lambda^{1/2})(\operatorname{Cov}\tilde{S})(\Lambda^{1/2} \otimes \Lambda^{1/2}) = (\gamma - \beta)\Lambda \otimes \Lambda + \beta\Lambda \square \Lambda$. For (iv), that $\mathcal{L}(S_{11}) = IW(\Lambda_{11}, q, \nu)$, take $A = (I_q\ 0)$ in part (i). To show $\mathcal{L}(S_{22\cdot 1}^{-1}) = W(\Lambda_{22\cdot 1}^{-1}, r, \nu + q + r - 1)$, use Proposition 8.8 on S^{-1}, which is $W(\Lambda^{-1}, p, \nu + p - 1)$.
8. For (i), let $p_1(x)p_2(s)$ denote the joint density of X and S with respect to the measure $dx\, ds/|s|^{(p+1)/2}$. Setting $T = XS^{-1/2}$ and $V = S$, the joint density of T and V is $p_1(tv^{1/2})p_2(v)|v|^{r/2}$ with respect to $dt\, dv/|v|^{(p+1)/2}$—the Jacobian of $x \to tv^{1/2}$ is $|v|^{r/2}$—see Proposition 5.10. Now, integrate out v to get the claimed density. That $\mathcal{L}(T) = \mathcal{L}(\Gamma T\Delta')$ is clear from the form of the density (also from (ii) below). Use Proposition 2.19 to show $\operatorname{Cov}(T) = c_1 I_r \otimes I_p$. Part (ii) follows by integrating out v from the conditional density of T to obtain the marginal density of T as given in (i). For (iii) represent T as: T given V is $N(0, I_r \otimes V)$ where V is $IW(I_p, p, \nu)$. Thus T_{11} given V is $N(0, I_k \otimes V_{11})$ where V_{11} is the $q \times q$ upper left-hand corner of V. Since $\mathcal{L}(V_{11}) = IW(I_q, q, \nu)$, the claimed result follows from (ii).
9. With $V = S_2^{-1/2}S_1S_2^{-1/2}$ and $S = S_2^{-1}$, the conditional distribution of V given S is $W(S, p, m)$ and $\mathcal{L}(S) = IW(I_p, p, \nu)$. Since V is unconditionally $F(m, \nu, I_p)$, (i) follows. For (ii), $\mathcal{L}(T) = T(\nu, I_r, I_p)$ means that $\mathcal{L}(T) = \mathcal{L}(XS^{1/2})$ where $\mathcal{L}(X) = N(0, I_r \otimes I_p)$ and $\mathcal{L}(S) = IW(I_p, p, \nu)$. Thus $\mathcal{L}(T'T) = \mathcal{L}(S^{1/2}X'XS^{1/2})$. Since $\mathcal{L}(X'X) = W(I_p, p, r)$, (ii) follows by definition of $F(r, \nu, I_p)$. For (iii), write $F = T'T$ where $\mathcal{L}(T) = T(\nu, I_r, I_p)$, which has the density given in (i) of Problem 8. Since $r \geqslant p$, Proposition 7.6 is directly applicable to yield the density of F. To establish (iv), first note that $\mathcal{L}(F) = \mathcal{L}(\Gamma F\Gamma')$

for all $\Gamma \in \mathcal{O}_p$. Using Example 7.16, F has the same distributions as $\psi D \psi'$ where ψ is uniform on $\mathcal{O}_p$ and is independent of the diagonal matrix D whose diagonal elements $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ are distributed as the eigenvalues of F. Thus $\lambda_1, \ldots, \lambda_p$ are distributed as the eigenvalues of $S_2^{-1}S_1$ where S_1 is $W(I_p, p, r)$ and S_2^{-1} is $IW(I_p, p, \nu)$. Hence $\mathcal{L}(F^{-1}) = \mathcal{L}(\psi D^{-1}\psi') = \mathcal{L}(\psi \tilde{D} \psi')$ where the diagonal elements of $\tilde{D}$, say $\lambda_p^{-1} \geqslant \cdots \geqslant \lambda_1^{-1}$, are the eigenvalues of $S_1^{-1}S_2$. Since S_2 is $W(I_p, p, \nu + p - 1)$, it follows that $\psi \tilde{D} \psi'$ has the same distribution as an $F(\nu + p - 1, r - p + 1, I_p)$ matrix by just repeating the orthogonal invariance argument given above. (v) is established by writing $F = T'T$ as in (ii) and partitioning T as $T_1 : r \times q$ and $T_2 : r \times (p - q)$ so

$$T'T = \begin{pmatrix} T_1'T_1 & T_1'T_2 \\ T_2'T_1 & T_2'T_2 \end{pmatrix}.$$

Since $\mathcal{L}(T_1) = T(\nu, I_r, I_q)$ and $F_{11} = T_1'T_1$, (ii) implies that $\mathcal{L}(F_{11}) = F(r, \nu, I_q)$. (vi) can be established by deriving the density of $XS^{-1}X'$ directly and using (iii), but an alternative argument is more instructive. First, apply Proposition 7.4 to X' and write $X = V^{1/2}\psi'$ where $V \in \mathcal{S}_r^+$, $V = XX'$ is $W(I_r, r, p)$ and is independent of $\psi : p \times r$, which is uniform on $\mathcal{F}_{r,p}$. Then $XS^{-1}X' = V^{1/2}W^{-1}V^{1/2}$ where $W = (\psi' S^{-1} \psi)^{-1}$ and is independent of V. Proposition 8.1 implies that $\mathcal{L}(W) = W(I_r, r, m - p + r)$. Thus $\mathcal{L}(W^{-1}) = IW(I_r, r, m - p + 1)$. Now, use the orthogonal invariance of the distribution of $XS^{-1}X'$ to conclude that $\mathcal{L}(XS^{-1}X') = \mathcal{L}(\Gamma D \Gamma')$ where Γ and D are independent, Γ is uniform on $\mathcal{O}_r$, and the diagonal elements of D are distributed as the ordered eigenvalues of $W^{-1}V$. As in the proof of (iv), conclude that $\mathcal{L}(\Gamma D \Gamma') = F(p, m - p + 1, I_r)$.

10. The function $S \to S^{1/2}$ on $\mathcal{S}_p^+$ to $\mathcal{S}_p^+$ satisfies $(\Gamma S \Gamma')^{1/2} = \Gamma S^{1/2} \Gamma'$ for $\Gamma \in \mathcal{O}_p$. With $B(S_1, S_2) = (S_1 + S_2)^{-1/2} S_1 (S_1 + S_2)^{-1/2}$, it follows that $B(\Gamma S_1 \Gamma', \Gamma S_2 \Gamma') = \Gamma B(S_1, S_2)\Gamma'$. Since $\mathcal{L}(\Gamma S_i \Gamma') = \mathcal{L}(S_i)$, $i = 1, 2$, and S_1 and S_2 are independent, the above implies that $\mathcal{L}(B) = \mathcal{L}(\Gamma B \Gamma')$ for $\Gamma \in \mathcal{O}_p$. The rest of (i) is clear from Example 7.16. For (ii), let $B_1 = S_1^{1/2}(S_1 + S_2)^{-1}S_1^{1/2}$ so $\mathcal{L}(B_1) = \mathcal{L}(\Gamma B_1 \Gamma')$ for $\Gamma \in \mathcal{O}_p$. Thus $\mathcal{L}(B_1) = \mathcal{L}(\psi D \psi')$ where ψ and D are independent, ψ is uniform on $\mathcal{O}_p$. Also, the diagonal elements of D, say $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$, are distributed as the ordered eigenvalues of $S_1(S_1 + S_2)^{-1}$ so B_1 is $B(m_1, m_2, I_p)$. (iii) is easy using (i) and (ii) and the fact that $F(I + F)^{-1}$ is symmetric. For (iv), let $B = X(S + X'X)^{-1}X'$ and observe that $\mathcal{L}(B) = \mathcal{L}(\Gamma B \Gamma')$, $\Gamma \in \mathcal{O}_p$. Since $m \geqslant p$, S^{-1} exists so $B = XS^{-1/2}(I_p + S^{-1/2}X'XS^{-1/2})^{-1}S^{-1/2}X'$. Hence $T = XS^{-1/2}$ is $T(m - p + 1, I_r, I_p)$. Thus $\mathcal{L}(B) = \mathcal{L}(\psi D \psi')$ where ψ is uniform on $\mathcal{O}_r$ and

is independent of D. The diagonal elements of D, say $\lambda_1,\ldots,\lambda_r$, are the eigenvalues of $T(I_p + T'T)^{-1}T'$. These are the same as the eigenvalues of $TT'(I_r + TT')^{-1}$ (use the singular value decomposition for T). But $\mathcal{L}(TT') = \mathcal{L}(XS^{-1}X') = F(p, m-p+1, I_r)$ by Problem 9 (vi). Now use (iii) above and the orthogonal invariance of $\mathcal{L}(B)$. (v) is trivial.

CHAPTER 9

1. Let B have rows $\nu_1',\ldots,\nu_k'$ and form X in the usual way (see Example 4.3) so $\mathcal{E}X = ZB$ with an appropriate $Z: n \times k$. Let $R: 1 \times k$ have entries $a_1,\ldots,a_k$. Then $RB = \Sigma_1^k a_i\mu_i'$ and H_0 holds iff $RB = 0$. Now apply the results in Section 9.1.
2. For (i), just do the algebra. For (ii), apply (i) with $S_1 = (Y - X\hat{B})'(Y - X\hat{B})$ and $S_2 = (X(B-\hat{B}))'(X(B-\hat{B}))$, so $\phi(S_1) \leqslant \phi(S_1 + S_2)$ for every B. Since $A \geqslant 0$, $\operatorname{tr} A(S_1 + S_2) = \operatorname{tr} AS_1 + \operatorname{tr} AS_2 \geqslant \operatorname{tr} AS_1$ since $\operatorname{tr} AS_2 \geqslant 0$ as $S_2 \geqslant 0$. To show $\det(A+S)$ is nondecreasing in $S \geqslant 0$, First note that $A + S_1 \leqslant A + S_1 + S_2$ in the sense of positive definiteness as $S_2 \geqslant 0$. Thus the ordered eigenvalues of $(A + S_1 + S_2)$, say $\lambda_1,\ldots,\lambda_p$, satisfy $\lambda_i \geqslant \mu_i$, $i = 1,\ldots,p$, where $\mu_1,\ldots,\mu_p$ are the ordered eigenvalues of $A + S_1$. Thus $\det(A + S_1 + S_2) \geqslant \det(A + S_1)$. This same argument solves (iv).
3. Since $\mathcal{L}(E\psi'A') = \mathcal{L}(EA')$ for $\psi \in \mathcal{O}_p$, the distribution of EA' depends only on a maximal invariant under the action $A \to A\psi$ of ψ on Gl_p. This maximal invariant is AA'. (ii) is clear and (iii) follows since the reduction to canonical form is achieved via an orthogonal transformation $\tilde{Y} = \Gamma Y$ where $\Gamma \in \mathcal{O}_n$. Thus $\tilde{Y} = \Gamma\mu + \Gamma EA'$. Γ is chosen so $\Gamma\mu$ has the claimed form and H_0 is $\tilde{B}_1 = 0$. Setting $\tilde{E} = \Gamma E$, the model has the claimed form and $\mathcal{L}(E) = \mathcal{L}(\tilde{E})$ by assumption. The arguments given in Section 9.1 show that the testing problem is invariant and a maximal invariant is the vector of the t largest eigenvalues of $Y_1(Y_3'Y_3)^{-1}Y_1'$. Under H_0, $Y_1 = E_1A'$, $Y_3 = E_3A'$ so $Y_1(Y_3'Y_3)^{-1}Y_1' = E_1(E_3'E_3)^{-1}E_1' \equiv W$. When $\mathcal{L}(\Gamma E) = \mathcal{L}(E)$ for all $\Gamma \in \mathcal{O}_n$, write $E = \psi U$ according to Proposition 7.3 where ψ and U are independent and ψ is uniform on $\mathcal{F}_{p,n}$. Partitioning ψ as E is partitioned, $E_i = \psi_i U$, $i = 1,2,3$, so $W = \psi_1 U((\psi_3 U)'\psi_3 U)^{-1}U'\psi_1' = \psi_1(\psi_3'\psi_3)^{-1}\psi_1'$. The rest is obvious as the distribution of W depends only on the distribution of ψ.
4. Use the independence of Y_1 and Y_3 and the fact that $\mathcal{E}(Y_3'Y_3)^{-1} = (m - p - 1)^{-1}\Sigma^{-1}$.

5. Let $\Gamma \in \mathcal{O}_2$ be given by

$$\Gamma = (\sqrt{2})^{-1}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

and set $\tilde{Y} = Y\Gamma$. Then $\mathcal{L}(\tilde{Y}) = N(ZB\Gamma, I_n \otimes \Gamma'\Sigma\Gamma)$. Now, let $B\Gamma$ have columns β_1 and β_2. Then H_0 is that $\beta_1 = 0$. Also $\Gamma'\Sigma\Gamma$ is diagonal with unknown diagonal elements. The results of Section 9.2 apply directly to yield the likelihood ratio test. A standard invariance argument shows the test is UMP invariant.

6. For (i), look at the i, j elements of the equation for Y. To show $M_2 \perp M_3$, compute as follows: $\langle \alpha u_2', u_1\beta' \rangle = \operatorname{tr} \alpha u_2' \beta u_1' = u_2' \beta u_1' \alpha = 0$ from the side conditions on α and β. The remaining relations $M_1 \perp M_2$ and $M_1 \perp M_2$ are verified similarly. For (iii) consider $(I_m \otimes A)(\mu u_1 u_2' + \alpha u_2' + u_1 \beta') = \mu u_1 (Au_2)' + \alpha(Au_2)' + u_1(A\beta)' = \mu\gamma u_1 u_2' + \gamma\alpha u_2' + \delta u_1 \beta' \in M$ where the relations $Pu_2 = u_2$ and $Q\beta = \beta$ when $u_2'\beta = 0$ have been used. This shows that M is invariant under each $I_m \otimes A$. It is now readily verified that $\hat{\mu} = \overline{Y}_{..}$, $\hat{\alpha}_i = \hat{Y}_{i.} - \overline{Y}_{..}$ and $\hat{\beta}_j = \overline{Y}_{.j} - \overline{Y}_{..}$. For (iv), first note that the subspace $\omega = \{x | x \in M, \alpha = 0\}$ defined by H_0 is invariant under each $I_m \otimes A$. Obviously, $\omega = M_1 \oplus M_3$. Consider the group whose elements are $g = (c, \Gamma, b)$ where c is a positive scalar, $b \in M_1 \oplus M_3$, and Γ is an orthogonal transformation with invariant subspaces M_2, $M_1 \oplus M_3$, and $M^\perp$. The testing problem is invariant under $x \to c\Gamma x + b$ and a maximal invariant is W (up to a set a measure zero). Since W has a noncentral F-distribution, the test that rejects for large values of W is UMP invariant.

7. (i) is clear. The column space of W is contained in the column space of Z and has dimension r. Let $x_1, \ldots, x_r, x_{r+1}, \ldots, x_k, x_{k+1}, \ldots, x_n$ be an orthonormal basis for R^n such that $\operatorname{span}\{x_1, \ldots, x_r\}$ = column space of W and $\operatorname{span}\{x_1, \ldots, x_k\}$ = column space of Z. Also, let $y_1, \ldots, y_p$ be any orthonormal basis for R^p. Then $\{x_i \square y_j | i = 1, \ldots, r, j = 1, \ldots, p\}$ is a basis for $\mathfrak{R}(P_W \otimes I_p)$, which has dimension rp. Obviously, $\mathfrak{R}(P_W \otimes I_p) \subseteq M$. Consider $x \in \omega$ so $x = ZB$ with $RB = 0$. Thus $(P_W \otimes I_p)x = P_W ZB = W(W'W)^{-1}W'ZB = W(W'W)^{-1}R(Z'Z)^{-1}(ZZ)B = W(W'W)^{-1}RB = 0$. Thus $\mathfrak{N}(P_W \otimes I_p) \supseteq \omega$, which implies $\mathfrak{R}(P_W \otimes I_p) \subseteq \omega^\perp$. Hence $\mathfrak{R}(P_W \otimes I_p) \subseteq M \cap \omega^\perp$. That $\dim \omega = (k - r)p$ can be shown by a reduction to canonical form as was done in Section 9.1. Since $\omega \subseteq M$, $\dim(M - \omega) = \dim M - \dim \omega = rp$, which entails $\mathfrak{R}(P_W \otimes I_p) = M - \omega$. Hence $P_Z \otimes I_p - P_W \otimes I_p$ is the orthogonal projection onto ω.

8. Use the fact that $\Gamma\Sigma\Gamma$ is diagonal with diagonal entries $\alpha_1, \alpha_2, \alpha_3, \alpha_3, \alpha_2$ (see Proposition 9.13 ff.) so the maximum likelihood estimators $\alpha_1, \alpha_2,$

and α_3 are easy to find—just transform the data by Γ. Let $\hat{D}$ have diagonal entries $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_3$, $\hat{\alpha}_2$ so $\hat{\Sigma} = \Gamma\hat{D}\Gamma$ gives the maximum likelihood estimators of σ^2, ρ_1, and ρ_2.

9. Do the problems in the complex domain first to show that if $Z_1, \ldots, Z_n$ are i.i.d. $\mathbb{C}N(0, 2H)$, then $\hat{H} = (1/2n)\Sigma_1^n Z_j Z_j^*$. But if $Z_j = U_j + iV_j$ and

$$X_j = \begin{pmatrix} U_j \\ V_j \end{pmatrix},$$

then $\hat{H} = (1/2n)\Sigma_1^n (U_j + iV_j)(U_j - iV_j)' = (1/2n)[(S_{11} + S_{22}) + i(S_{12} - S_{21})]$ so $\hat{\psi} = \{\hat{H}\}$. This gives the desired result.

10. Write $R = M(I_r\ 0)\Gamma$ where M is $r \times r$ of rank r and $\Gamma \in \mathcal{O}_p$. With $\delta = \Gamma\mu$, the null hypothesis is $(I_r\ 0)\delta = 0$. Now, transform the data by Γ and proceed with the analysis as in the first testing problem considered in Section 9.6.

11. First write $P_Z = P_1 + P_2$ where P_1 is the orthogonal projection onto e and P_2 is the orthogonal projection onto (column space of Z) $\cap$ $\{\text{span}\, e\}^\perp$. Thus $P_M = P_1 \otimes I_p + P_2 \otimes I_p$. Also, write $A(\rho) = \gamma P_1 + \delta Q_1$ where $\gamma = 1 + (n-1)\rho$, $\delta = 1 - \rho$, and $Q_1 = I_n - P_1$. The relations $P_1 P_2 = 0 = Q_1 P_1$ and $P_2 Q_1 = Q_1 P_2 = P_2$ show that M is invariant under $A(\rho) \otimes \Sigma$ for each value of ρ and Σ. Write $ZB = eb_1' + \Sigma_2^k z_j b_j'$ so $Q_1 Y$ is $N(\Sigma_2^k (Q_1 z_j) b_j', (Q_1 A(\rho) Q_1) \otimes \Sigma)$. Now, $Q_1 A(\rho) Q_1 = \delta Q_1$ so $Q_1 Y$ is $N(\beta_2^k (Q_1 z_j) b_j', \delta Q_1 \otimes \Sigma)$. Also, $P_1 Y$ is $N(eb_1', \gamma P_1 \otimes \Sigma)$. Since hypotheses of the form $\dot{R}\dot{B} = 0$ involve only $b_2, \ldots, b_p$, an invariance argument shows that invariant tests of H_0 will not involve $P_1 Y$—so just ignore $P_1 Y$. But the model for $Q_1 Y$ is of the MANOVA type; change coordinates so

$$Q_1 = \begin{pmatrix} I_{n-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Now, the null hypothesis is of the type discussed in Section 9.1.

CHAPTER 10

1. Part (i) is clear since the number of nonzero canonical correlations is always the rank of Σ_{12} in the partitioned covariance of $\{X, Y\}$. For (ii), write

$$\operatorname{Cov}\{\tilde{X}, \tilde{Y}\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where Σ_{12} has rank t, and $\Sigma_{11} > 0$, $\Sigma_{22} > 0$. First, consider the case when $q \leqslant r$, $\Sigma_{11} = I_q$, $\Sigma_{22} = I_r$, and

$$\Sigma_{12} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where $D > 0$ is $t \times t$ and diagonal. Set

$$A = \begin{pmatrix} D^{1/2} \\ 0 \end{pmatrix} : q \times t, \qquad B = \begin{pmatrix} D^{1/2} \\ 0 \end{pmatrix} : r \times t$$

so $AB' = \Sigma_{12}$. Now, set $\Lambda_{11} = I_q - AA'$, $\Lambda_{22} = I_r - BB'$, and the problem is solved for this case. The general case is solved by using Proposition 5.7 to reduce the problem to the case above.

2. That $\Sigma_{12} = \delta e_1 e_2'$ for some $\delta \in R^1$ is clear, and hence Σ_{12} has rank one–hence at most one nonzero canonical correlation. It is the square root of the largest eigenvalue of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \delta^2\Sigma_{11}^{-1}e_1e_2'\Sigma_{22}^{-1}e_2e_1'$. The only nonzero (possibly) eigenvalue is $\delta^2 e_1'\Sigma_{11}^{-1}e_1e_2'\Sigma_{22}^{-1}e_2$. To describe canonical coordinates, let

$$\tilde{v}_1 = \frac{\Sigma_{11}^{-1/2}e_1}{\|\Sigma_{11}^{-1/2}e_1\|}, \qquad \tilde{w}_1 = \frac{\Sigma_{22}^{-1/2}e_2}{\|\Sigma_{22}^{-1/2}e_2\|}$$

and then form orthonormal bases $\{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_q\}$ and $\{\tilde{w}_1, \ldots, \tilde{w}_r\}$ for R^q and R^r. Now, set $v_i = \Sigma_{11}^{-1/2}\tilde{v}_i$, $w_j = \Sigma_{22}^{-1/2}\tilde{w}_j$ for $i = 1, \ldots, q$, $j = 1, \ldots, r$. Then verify that $X_i = v_i'X$ and $Y_j = w_j'Y$ form a set of canonical coordinates for X and Y.

3. Part (i) follows immediately from Proposition 10.4 and the form of the covariance for $\{X, Y\}$. That $\delta(B) = \operatorname{tr} A(I - Q(B))$ is clear and the minimization of $\delta(B)$ follows from Proposition 1.44. To describe $\hat{B}$, let $\psi : p \times t$ have columns $a_1, \ldots, a_t$ so $\psi'\psi = I_t$ and $\hat{Q} = \psi\psi'$. Then show directly that $\hat{B} = \psi'\Sigma^{-1/2}$ is the minimizer and $\hat{C}\hat{B}X = \Sigma^{1/2}\hat{Q}\Sigma^{-1/2}X$ is the best predictor. (iii) is an immediate application of (ii).

4. Part (i) is easy. For (ii), with $u_i = x_i - a_0$,

$$\Delta(M, a_0) = \sum_1^n \|x_i - (P(x_i - a_0) + a_0)\|^2 = \sum_1^n \|u_i - Pu_i\|^2$$

$$= \sum_1^n \|Qu_i\|^2 = \sum_1^n \operatorname{tr} Qu_iu_i' = \operatorname{tr} Q\sum_1^n u_iu_i' = \operatorname{tr} S(a_0)Q.$$

Since $S(a_0) = S(\bar{x}) + n(\bar{x} - a_0)(\bar{x} - a_0)'$, (ii) follows. (iii) is an application of Proposition 1.44.

6. Part (i) follows from the singular value decomposition: For (ii), $\{x \in \mathcal{L}_{p,n} | x = \psi C, C \in \mathcal{L}_{p,k}\}$ is a linear subspace of $\mathcal{L}_{p,n}$ and the orthogonal projection onto this subspace is $(\psi\psi') \otimes I_p$. Thus the closest point to A is $((\psi\psi') \otimes I)A = \psi\psi' A$, and the C that achieves the minimum is $\hat{C} = \psi' A$. For $B \in \mathfrak{B}_k$, write $B = \psi C$ as in (i). Then

$$\|A - B\|^2 \geq \inf_{\psi} \inf_{C} \|A - \psi C\|^2 = \inf_{\psi} \|A - \psi\psi' A\|^2 = \inf_{Q} \|AQ\|^2.$$

The last equality follows as each ψ determines a Q and conversely. Since $\|AQ\|^2 = \operatorname{tr} AQ(AQ)' = \operatorname{tr} AQ^2 A' = \operatorname{tr} QAA'$,

$$\|A - B\|^2 \geq \inf_{Q} \operatorname{tr} QAA'.$$

Writing $A = \sum_1^p \lambda_i u_i v_i'$ (the singular value decomposition for A), $AA' = \sum_1^p \lambda_i u_i u_i'$ is a spectral decomposition for AA'. Using Proposition 1.44, it follows easily that

$$\inf_{Q} \operatorname{tr} QAA' = \sum_{k+1}^{p} \lambda_i^2.$$

That $\hat{B}$ achieves the infimum is a routine calculation.

7. From Proposition 10.8, the density of W is

$$h(w|\theta) = \int_0^\infty p_{n-2}(w|\theta u^{1/2}) f(u)\, du$$

where p_{n-2} is the density of a noncentral t distribution and f is the density of a χ^2_{n-1} distribution. For $\theta > 0$, set $v = \theta u^{1/2}$ so

$$h(w|\theta) = \frac{2}{\theta^2} \int_0^\infty p_{n-2}(w|v) f\left(\frac{v^2}{\theta^2}\right) v\, dv.$$

Since $p_{n-2}(w|v)$ has a monotone likelihood ratio in w and v and $f(v^2/\theta^2)$ has a monotone likelihood ratio in v and θ, Karlin's Lemma implies that $h(w|\theta)$ has a monotone likelihood ratio. For $\theta < 0$, set $v = \theta u^{-1/2}$, change variables, and use Karlin's Lemma again. The last assertion is clear.

8. For U_2 fixed, the conditional distribution of W given U_2 can be described as the ratio of two independent random variables—the numerator has a χ^2_{r+2K} distribution (given K) and K is Poisson with parameter $\Delta/2$ where $\Delta = \rho^2(1-\rho^2)^{-1}U_2$ and the denominator is χ^2_{n-r-1}. Hence, given U_2, this ratio is $\mathcal{F}_{r+2K,\,n-r-1}$ with K described above, so the conditional density of W is

$$f_1(w|\rho, U_2) = \sum_{k=0}^{\infty} f_{r+2k,\,n-r-1}(w)\psi\left(k|\frac{\Delta}{2}\right)$$

where $\psi(\cdot|\Delta/2)$ is the Poisson probability function. Integrating out U_2 gives the unconditional density of W (at ρ). Thus it must be shown that $\mathcal{E}_{U_2}\psi(k|\Delta/2) = h(k|\rho)$—this is a calculation. That $f(\cdot|\rho)$ has a monotone likelihood ratio is a direct application of Karlin's Lemma.

9. Let M be the range of P. Each $R \in \mathcal{P}_s$ can be represented as $R = \psi\psi'$ where ψ is $n \times s$, $\psi'\psi = I_s$, and $P\psi = 0$. In other words, R corresponds to orthonormal vectors $\psi_1, \ldots, \psi_s$ (the columns of ψ) and these vectors are in $M^{\perp}$ (of course, these vectors are not unique). But given any two such sets—say $\psi_1, \ldots, \psi_s$ and $\delta_1, \ldots, \delta_s$, there is a $\Gamma \in \mathcal{O}(P)$ such that $\Gamma\psi_i = \delta_i$, $i = 1, \ldots, s$. This shows $\mathcal{O}(P)$ is compact and acts transitively on $\mathcal{P}_s$, so there is a unique $\mathcal{O}(P)$ invariant probability distribution on $\mathcal{P}_s$. For (iii), $\Delta R_0 \Delta'$ has an $\mathcal{O}(P)$ invariant distribution on $\mathcal{P}_s$—uniqueness does the rest.

10. For (i), use Proposition 7.3 to write $Z = \psi U$ with probability one where ψ and U are independent, ψ is uniform on $\mathcal{F}_{p,n}$, and $U \in G_U^+$. Thus with probability one, rank(QZ) = rank$(Q\psi)$. Let $S \geqslant 0$ be independent of ψ with $\mathcal{L}(S^2) = W(I_p, p, n)$ so S has rank p with probability one. Thus rank$(Q\psi)$ = rank$(Q\psi S)$ with probability one. But ψS is $N(0, I_n \otimes I_p)$, which implies that $Q\psi S$ has rank p. Part (ii) is a direct application of Problem 9.

12. That ψ is uniform follows from the uniformity of Γ on $\mathcal{O}_n$. For (ii), $\mathcal{L}(\psi) = \mathcal{L}(Z(Z'Z)^{-1/2})$ and $\Delta = (I_k\ 0)\psi$ implies that $\mathcal{L}(\psi) = \mathcal{L}(X(X'X + Y'Y)^{-1})$. (iii) is immediate from Problem 11, and (iv) is an application of Proposition 7.6. For (v), it suffices to show that $\int f(x)P_1(dx) = \int f(x)P_2(dx)$ for all bounded measurable f. The invariance of P_i implies that for $i = 1,2$, $\int f(x)P_i(dx) = \int f(gx)P_i(dx)$, $g \in G$. Let ν be uniform probability measure on G and integrate the above to get $\int f(x)P_i(dx) = \int(\int_G f(gx)\nu(dg))P_i(dx)$. But the function $x \to \int_G f(gx)\nu(dg)$ is G-invariant and so can be written $\hat{f}(\tau(x))$ as τ is a maximal invariant. Since $P_1(\tau^{-1}(C)) = P_2(\tau^{-1}(C))$ for all measurable C, we have $\int k(\tau(x))P_1(dx) = \int k(\tau(x))P_2(dx)$ for all bounded

measurable k. Putting things together, we have $\int f(x)P_1(dx) = \int \hat{f}(\tau(x))P_1(dx) = \int \hat{f}(\tau(x))P_2(dx) = \int f(x)P_2(dx)$ so $P_1 = P_2$. Part (vi) is immediate from (v).

13. For (i), argue as in Example 4.4:

$$\begin{aligned}
&\operatorname{tr}(Z - TB)\Sigma^{-1}(Z - TB)' \\
&\quad = \operatorname{tr}\big(Z - T\hat{B} + T(\hat{B} - B)\big)\Sigma^{-1}\big(Z - T\hat{B} + T(\hat{B} - B)\big)' \\
&\quad = \operatorname{tr}\big(QZ + T(\hat{B} - B)\big)\Sigma^{-1}\big(QZ + T(\hat{B} - B)\big)' \\
&\quad = \operatorname{tr}(QZ)\Sigma^{-1}(QZ)' + \operatorname{tr} T(\hat{B} - B)\Sigma^{-1}(\hat{B} - B)'T' \\
&\quad \geqslant \operatorname{tr}(QZ)\Sigma^{-1}(QZ)' = \operatorname{tr} Z'QZ\Sigma^{-1}.
\end{aligned}$$

The third equality follows from the relation $QT = 0$ as in the normal case. Since h is nonincreasing, this shows that for each $\Sigma > 0$,

$$\sup_B f(Z|B, \Sigma) = f(Z|\hat{B}, \Sigma)$$

and it is obvious that $f(Z|\hat{B}, \Sigma) = |\Sigma|^{-n/2}h(\operatorname{tr} S\Sigma^{-1})$. For (ii), first note that $S > 0$ with probability one. Then, for $S > 0$,

$$\begin{aligned}
\sup_{H_1 \cup H_0} f(Z|B, \Sigma) &= \sup_{\Sigma > 0} f(Z|\hat{B}, \Sigma) \\
&= \sup_{\Sigma > 0} |\Sigma|^{-n/2}h(\operatorname{tr} S\Sigma^{-1}) \\
&= |S|^{-n/2} \sup_{C > 0} |C|^{n/2}h(\operatorname{tr} C).
\end{aligned}$$

Under H_0, we have

$$\begin{aligned}
&\sup_{H_0} f(Z|B, \Sigma) \\
&\quad = \sup_{\Sigma_{ii} > 0,\, i = 1,2} |\Sigma_{11}|^{-n/2}|\Sigma_{22}|^{-n/2}h\big(\operatorname{tr} \Sigma_{11}^{-1}S_{11} + \operatorname{tr} \Sigma_{22}^{-1}S_{22}\big) \\
&\quad = |S_{11}|^{-n/2}|S_{22}|^{-n/2} \sup_{C_{ii} > 0,\, i = 1,2} |C_{11}|^{n/2}|C_{22}|^{n/2}h(\operatorname{tr} C_{11} + \operatorname{tr} C_{22}).
\end{aligned}$$

This latter sup is bounded above by

$$\sup_{C > 0} |C|^{n/2}h(\operatorname{tr} C) \equiv k,$$

which is finite by assumption. Hence the likelihood ratio test rejects for small values of $k_1|S_{11}|^{-n/2}|S_{22}|^{-n/2}|S|^{n/2}$, which is equivalent to rejecting for small values of $\Lambda(Z)$. The identity of part (iii) follows from the equations relating the blocks of Σ to the blocks of Σ^{-1}. Partition B into $B_1 : k \times q$ and $B_2 : k \times r$ so $\mathcal{E}X = TB_1$ and $\mathcal{E}Y = TB_2$. Apply the identity with $U = X - TB_1$ and $V = Y - TB_2$ to give

$$\begin{aligned} f(Z|B, \Sigma) = {} & |\Sigma_{11}|^{-n/2}|\Sigma_{22\cdot 1}|^{-n/2} \\ & \times h\Big[\operatorname{tr}\big(Y - TB_2 - (X - TB_1)\Sigma_{11}^{-1}\Sigma_{12}\big) \\ & \times \Sigma_{22\cdot 1}^{-1}\big(Y - TB_2 - (X - TB_1)\Sigma_{11}^{-1}\Sigma_{12}\big)' \\ & + \operatorname{tr}(X - TB_1)\Sigma_{11}^{-1}(X - TB_1)'\Big]. \end{aligned}$$

Using the notation of Section 10.5, write

$$\begin{aligned} f(X, Y|B, \Sigma) = {} & |\Sigma_{11}|^{-n/2}|\Sigma_{22\cdot 1}|^{-n/2} \\ & \times h\big[\operatorname{tr}(Y - WC)\Sigma_{22\cdot 1}^{-1}(Y - WC)' \\ & + \operatorname{tr}(X - TB_1)\Sigma_{11}^{-1}(X - TB_1)'\big]. \end{aligned}$$

Hence the conditional density of Y given X is

$$\begin{aligned} & f_1(Y|C, B_1, \Sigma_{11}, \Sigma_{22\cdot 1}, X) \\ & \quad = |\Sigma_{22\cdot 1}|^{-n/2} h\big(\operatorname{tr}(Y - WC)\Sigma_{22\cdot 1}^{-1}(Y - WC)' + \eta\big)\phi(\eta) \end{aligned}$$

where $\eta = \operatorname{tr}(X - TB_1)\Sigma_{11}^{-1}(X - TB_1)$ and $(\phi(\eta))^{-1} = \int_{\mathcal{L}_{r,n}} h(\operatorname{tr} uu' + \eta)\,du$. For (iv), argue as in (ii) and use the identities established in Proposition 10.17. Part (v) is easy, given the results of (iv)—just note that the sup over Σ_{11} and B_1 is equal to the sup over $\eta > 0$. Part (vi) is interesting—Proposition 10.13 is not applicable. Fix X, B_1, and Σ_{11} and note that under H_0, the conditional density of Y is

$$\begin{aligned} & f_2(Y|C_2, \Sigma_{22\cdot 1}, \eta) \\ & \quad = |\Sigma_{22\cdot 1}|^{-n/2} h\big(\operatorname{tr}(Y - TC_2)\Sigma_{22\cdot 1}^{-1}(Y - TC_2) + \eta\big)\phi(\eta). \end{aligned}$$

This shows that Y has the same distribution (conditionally) as $\tilde{Y} =$

$TC_2 + E\Sigma_{22\cdot 1}^{1/2}$ where $E \in \mathcal{L}_{r,n}$ has density $h(\operatorname{tr} EE' + \eta)\phi(\eta)$. Note that $\mathcal{L}(\Gamma E \Delta) = \mathcal{L}(E)$ for all $\Gamma \in \mathcal{O}_n$ and $\Delta \in \mathcal{O}_r$. Let $t = \min(q, r)$ and, given any $n \times n$ matrix A with real eigenvalues, let $\lambda(A)$ be the vector of the t largest eigenvalues of A. Thus the squares of the sample canonical correlations are the elements of the vector $\lambda(R_Y R_X)$ where $R_Y = (QY)(Y'QY)^{-1}(QY)$, $R_X = QX(X'QX)^{-1}QX$, since

$$S = \begin{pmatrix} X'QX & X'QY \\ Y'QX & Y'QY \end{pmatrix}.$$

(You may want to look at the discussion preceding Proposition 10.5.) Now, we use Problem 9 and the notation there—$P = I - Q$. First, $R_Y \in \mathcal{P}_r$, $R_X \in \mathcal{P}_q$, and $\mathcal{O}(P)$ acts transitively on $\mathcal{P}_r$ and $\mathcal{P}_q$. Under H_0 (and X fixed), $\mathcal{L}(QY) = \mathcal{L}(QE\Sigma_{22\cdot 1}^{1/2})$, which implies that $\mathcal{L}(\Gamma R_Y \Gamma') = \mathcal{L}(R_Y)$, $\Gamma \in \mathcal{O}(P)$. Hence R_Y is uniform on $\mathcal{P}_r$ for each X. Fix $R_0 \in \mathcal{R}_q$ and choose Γ_0 so that $\Gamma_0 R_0 \Gamma_0' = R_X$. Then, for each X,

$$\begin{aligned} \mathcal{L}(\lambda(R_Y R_0)) &= \mathcal{L}(\lambda(\Gamma_0 R_Y R_0 \Gamma_0')) = \mathcal{L}(\lambda(\Gamma_0 R_Y \Gamma_0' \Gamma_0 R_0 \Gamma_0')) \\ &= \mathcal{L}(\lambda(\Gamma_0 R_Y \Gamma_0' R_X) = \mathcal{L}(\lambda(R_Y R_X)). \end{aligned}$$

This shows that for each X, $\lambda(R_Y R_X)$ has the same distribution as $\lambda(R_Y R_0)$ for R_0 fixed where R_Y is uniform on $\mathcal{P}_r$. Since the distribution of $\lambda(R_Y R_0)$ does not depend on X and agrees with what we get in the normal case, the solution is complete.