The sample mean vector and the covariance matrix can be partitioned in order to distinguish quantities corresponding to groups of variables. Thus,

$$\underset{(p\times 1)}{\bar{\mathbf{x}}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_q \\ \hdashline \bar{x}_{q+1} \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \hdashline \bar{\mathbf{x}}^{(2)} \end{bmatrix} \tag{2-46}$$

and

$$\underset{(p\times p)}{\mathbf{S}_n} = \left[\begin{array}{ccc:ccc} s_{11} & \cdots & s_{1q} & s_{1,q+1} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{q1} & \cdots & s_{qq} & s_{q,q+1} & \cdots & s_{qp} \\ \hdashline s_{q+1,1} & \cdots & s_{q+1,q} & s_{q+1,q+1} & \cdots & s_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pq} & s_{p,q+1} & \cdots & s_{pp} \end{array}\right]$$

$$= \begin{array}{c} \\ q \\ p-q \end{array}\overset{\quad q \qquad p-q}{\left[\begin{array}{c:c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hdashline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right]} \tag{2-47}$$

where $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$ are the sample mean vectors constructed from observations $\mathbf{x}^{(1)} = [x_1, \ldots, x_q]'$ and $\mathbf{x}^{(2)} = [x_{q+1}, \ldots, x_p]'$, respectively; $\mathbf{S}_{11}$ is the sample covariance matrix computed from observations $\mathbf{x}^{(1)}$; $\mathbf{S}_{22}$ is the sample covariance matrix computed from observations $\mathbf{x}^{(2)}$; and $\mathbf{S}_{12} = \mathbf{S}'_{21}$ is the sample covariance matrix for elements of $\mathbf{x}^{(1)}$ and elements of $\mathbf{x}^{(2)}$.

2.7 Matrix Inequalities and Maximization

Maximization principles play an important role in several multivariate techniques. Linear discriminant analysis, for example, is concerned with allocating observations to predetermined groups. The allocation rule is often a linear function of measurements that *maximizes* the separation between groups relative to their within-group variability. As another example, principal components are linear combinations of measurements with *maximum* variability.

The matrix inequalities presented in this section will easily allow us to derive certain maximization results, which will be referenced in later chapters.

Cauchy-Schwarz Inequality. Let $\mathbf{b}$ and $\mathbf{d}$ be *any* two $p \times 1$ vectors. Then

$$(\mathbf{b}'\mathbf{d})^2 \le (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d}) \tag{2-48}$$

with equality if and only if $\mathbf{b} = c\,\mathbf{d}$ (or $\mathbf{d} = c\,\mathbf{b}$) for some constant c.

Proof. The inequality is obvious if either $\mathbf{b} = \mathbf{0}$ or $\mathbf{d} = \mathbf{0}$. Excluding this possibility, consider the vector $\mathbf{b} - x\,\mathbf{d}$, where x is an arbitrary scalar. Since the length of $\mathbf{b} - x\mathbf{d}$ is positive for $\mathbf{b} - x\mathbf{d} \ne \mathbf{0}$, in this case

$$\begin{aligned} 0 < (\mathbf{b} - x\mathbf{d})'(\mathbf{b} - x\mathbf{d}) &= \mathbf{b}'\mathbf{b} - x\mathbf{d}'\mathbf{b} - \mathbf{b}'(x\mathbf{d}) + x^2\mathbf{d}'\mathbf{d} \\ &= \mathbf{b}'\mathbf{b} - 2x(\mathbf{b}'\mathbf{d}) + x^2(\mathbf{d}'\mathbf{d}) \end{aligned}$$

The last expression is quadratic in x. If we complete the square by adding and subtracting the scalar $(\mathbf{b}'\mathbf{d})^2/\mathbf{d}'\mathbf{d}$, we get

$$\begin{aligned} 0 < \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} + \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} - 2x(\mathbf{b}'\mathbf{d}) + x^2(\mathbf{d}'\mathbf{d}) \\ = \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} + (\mathbf{d}'\mathbf{d})\left(x - \frac{\mathbf{b}'\mathbf{d}}{\mathbf{d}'\mathbf{d}}\right)^2 \end{aligned}$$

The term in brackets is zero if we choose $x = \mathbf{b}'\mathbf{d}/\mathbf{d}'\mathbf{d}$, so we conclude that

$$0 < \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}}$$

or $(\mathbf{b}'\mathbf{d})^2 < (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$ if $\mathbf{b} \ne x\mathbf{d}$ for some x.

Note that if $\mathbf{b} = c\mathbf{d}$, $0 = (\mathbf{b} - c\mathbf{d})'(\mathbf{b} - c\mathbf{d})$, and the same argument produces $(\mathbf{b}'\mathbf{d})^2 = (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$. ■

A simple, but important, extension of the Cauchy-Schwarz inequality follows directly.

Extended Cauchy-Schwarz Inequality. Let $\underset{(p\times 1)}{\mathbf{b}}$ and $\underset{(p\times 1)}{\mathbf{d}}$ be any two vectors, and let $\underset{(p\times p)}{\mathbf{B}}$ be a positive definite matrix. Then

$$(\mathbf{b}'\mathbf{d})^2 \le (\mathbf{b}'\mathbf{B}\,\mathbf{b})(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d}) \tag{2-49}$$

with equality if and only if $\mathbf{b} = c\,\mathbf{B}^{-1}\mathbf{d}$ (or $\mathbf{d} = c\,\mathbf{B}\,\mathbf{b}$) for some constant c.

Proof. The inequality is obvious when $\mathbf{b} = \mathbf{0}$ or $\mathbf{d} = \mathbf{0}$. For cases other than these, consider the square-root matrix $\mathbf{B}^{1/2}$ defined in terms of its eigenvalues λ_i and the normalized eigenvectors $\mathbf{e}_i$ as $\mathbf{B}^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i}\,\mathbf{e}_i\mathbf{e}'_i$. If we set [see also (2-22)]

$$\mathbf{B}^{-1/2} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}}\,\mathbf{e}_i\mathbf{e}'_i$$

it follows that

$$\mathbf{b}'\mathbf{d} = \mathbf{b}'\mathbf{I}\mathbf{d} = \mathbf{b}'\mathbf{B}^{1/2}\mathbf{B}^{-1/2}\mathbf{d} = (\mathbf{B}^{1/2}\mathbf{b})'(\mathbf{B}^{-1/2}\mathbf{d})$$

and the proof is completed by applying the Cauchy-Schwarz inequality to the vectors $(\mathbf{B}^{1/2}\mathbf{b})$ and $(\mathbf{B}^{-1/2}\mathbf{d})$. ■

The extended Cauchy-Schwarz inequality gives rise to the following maximization result.

Maximization Lemma. Let $\underset{(p\times p)}{\mathbf{B}}$ be positive definite and $\underset{(p\times 1)}{\mathbf{d}}$ be a given vector. Then, for an arbitrary nonzero vector $\underset{(p\times 1)}{\mathbf{x}}$,

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d} \tag{2-50}$$

with the maximum attained when $\underset{(p\times 1)}{\mathbf{x}} = c\underset{(p\times p)}{\mathbf{B}^{-1}}\underset{(p\times 1)}{\mathbf{d}}$ for any constant $c \neq 0$.

Proof. By the extended Cauchy-Schwarz inequality, $(\mathbf{x}'\mathbf{d})^2 \le (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d})$. Because $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{B}$ is positive definite, $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$. Dividing both sides of the inequality by the positive scalar $\mathbf{x}'\mathbf{B}\mathbf{x}$ yields the upper bound

$$\frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} \le \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$

Taking the maximum over $\mathbf{x}$ gives Equation (2-50) because the bound is attained for $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$. ∎

A final maximization result will provide us with an interpretation of eigenvalues.

Maximization of Quadratic Forms for Points on the Unit Sphere. Let $\underset{(p\times p)}{\mathbf{B}}$ be a positive definite matrix with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_1)$$
$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_p) \tag{2-51}$$

Moreover,

$$\max_{\mathbf{x}\perp\mathbf{e}_1,\ldots,\mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1} \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_{k+1}, k = 1, 2, \ldots, p-1) \tag{2-52}$$

where the symbol $\perp$ is read "is perpendicular to."

Proof. Let $\underset{(p\times p)}{\mathbf{P}}$ be the orthogonal matrix whose columns are the eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ and Λ be the diagonal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ along the main diagonal. Let $\mathbf{B}^{1/2} = \mathbf{P}\Lambda^{1/2}\mathbf{P}'$ [see (2-22)] and $\underset{(p\times 1)}{\mathbf{y}} = \underset{(p\times p)}{\mathbf{P}'}\underset{(p\times 1)}{\mathbf{x}}$.

Consequently, $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{y} \neq \mathbf{0}$. Thus,

$$\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}}{\mathbf{x}'\underbrace{\mathbf{P}\mathbf{P}'}_{\underset{(p\times p)}{\mathbf{I}}}\mathbf{x}} = \frac{\mathbf{x}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{x}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\Lambda\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$
$$= \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} \le \lambda_1 \frac{\sum_{i=1}^{p} y_i^2}{\sum_{i=1}^{p} y_i^2} = \lambda_1 \tag{2-53}$$

Setting $\mathbf{x} = \mathbf{e}_1$ gives

$$\mathbf{y} = \mathbf{P}'\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

since

$$\mathbf{e}_k'\mathbf{e}_1 = \begin{cases} 1, & k = 1 \\ 0, & k \neq 1 \end{cases}$$

For this choice of $\mathbf{x}$, we have $\mathbf{y}'\Lambda\mathbf{y}/\mathbf{y}'\mathbf{y} = \lambda_1/1 = \lambda_1$, or

$$\frac{\mathbf{e}_1'\mathbf{B}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{e}_1} = \mathbf{e}_1'\mathbf{B}\mathbf{e}_1 = \lambda_1 \tag{2-54}$$

A similar argument produces the second part of (2-51).

Now, $\mathbf{x} = \mathbf{P}\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \cdots + y_p\mathbf{e}_p$, so $\mathbf{x} \perp \mathbf{e}_1, \ldots, \mathbf{e}_k$ implies

$$0 = \mathbf{e}_i'\mathbf{x} = y_1\mathbf{e}_i'\mathbf{e}_1 + y_2\mathbf{e}_i'\mathbf{e}_2 + \cdots + y_p\mathbf{e}_i'\mathbf{e}_p = y_i, \quad i \le k$$

Therefore, for $\mathbf{x}$ perpendicular to the first k eigenvectors $\mathbf{e}_i$, the left-hand side of the inequality in (2-53) becomes

$$\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{i=k+1}^{p} \lambda_i y_i^2}{\sum_{i=k+1}^{p} y_i^2}$$

Taking $y_{k+1} = 1, y_{k+2} = \cdots = y_p = 0$ gives the asserted maximum. ∎

For a fixed $\mathbf{x}_0 \neq \mathbf{0}$, $\mathbf{x}_0'\mathbf{B}\mathbf{x}_0/\mathbf{x}_0'\mathbf{x}_0$ has the same value as $\mathbf{x}'\mathbf{B}\mathbf{x}$, where $\mathbf{x}' = \mathbf{x}_0'/\sqrt{\mathbf{x}_0'\mathbf{x}_0}$ is of unit length. Consequently, Equation (2-51) says that the largest eigenvalue, λ_1, is the maximum value of the quadratic form $\mathbf{x}'\mathbf{B}\mathbf{x}$ for all points $\mathbf{x}$ whose distance from the origin is unity. Similarly, λ_p is the smallest value of the quadratic form for all points $\mathbf{x}$ one unit from the origin. The largest and smallest eigenvalues thus represent extreme values of $\mathbf{x}'\mathbf{B}\mathbf{x}$ for points on the unit sphere. The "intermediate" eigenvalues of the $p \times p$ positive definite matrix $\mathbf{B}$ also have an interpretation as extreme values when $\mathbf{x}$ is further restricted to be perpendicular to the earlier choices.