Appendix to Lecture 3, Proofs

I provide a proof of theorem 3.15, because the mathematics involved has fairly wide application in economics. The theorem is an immediate consequence of two facts, that the characteristic values of a symmetric matrix are real and that for any symmetric bilinear form, there is an orthonormal basis such that the matrix representation of the bilinear form with respect to that basis is diagonal. I now demonstrate these facts.

Lemma 3.16: Every characteristic value of a symmetric matrix is real.

Proof: Let $A$ be an $N\times N$ symmetric matrix and let $\lambda = a + bi$ be a characteristic value with corresponding characteristic vector $x + yi$, where $x$ and $y$ belong to $R^N$ and $x + yi \neq 0$. I must show that $b = 0$. Since

$$Ax + iAy = A(x + iy) = \lambda(x + yi) = (a + bi)(x + yi) = ax - by + (bx + ay)i,$$

It follows that

$$Ax = ax - by$$

and

$$Ay = bx + ay.$$

Hence

$$ay^\top x - by^\top y = y^\top Ax = x^\top A^\top y = x^\top Ay = bx^\top x + ax^\top y = bx^\top x + ay^\top x,$$

where the third equation follows from the symmetry of $A$ and the second and last equations follow from the fact that $y^\top Ax$ and $ax^\top y$ are numbers and so equal their own transposes. Canceling $ay^\top x$ from the extreme left and right of these equations, we see that

$$-by^\top y = bx^\top x.$$

If $b \neq 0$, then $-y^\top y = x^\top x$, which implies that $x = y = 0$. This is impossible, since $x + yi$ is a characteristic vector and so is non-zero. Therefore $b = 0$.

The fact that the characteristic value $\lambda = a + bi$ is real in the above proof implies that the corresponding characteristic vector $x + yi$ may be chosen to be real, that is, we may assume that $y = 0$. This is so, because the matrix $A - \lambda I = A - aiI$ is singular and has real entries, so that a non-zero linear combination of its columns with real coefficients equals zero. These coefficients are the components of a characteristic vector of $A$ corresponding to the characteristic value $\lambda = a$.

The example given earlier of the matrix $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ shows that if a square matrix with
real entries is not symmetric, its characteristic values may be complex (i.e. not real) and none of its characteristic vectors may have real components.

I now turn to the diagonalization of symmetric matrices by means of orthonormal bases. Recall what happens to the matrix representation of the linear form \( f(x, y) = x^T Ay \) when we change the basis of the underlying space, \( \mathbb{R}^n \), over which \( f \) is defined. Let \( b^1, \ldots, b^n \) be a basis for \( \mathbb{R}^n \) and let \( B \) be the \( NxN \) matrix with columns \( b^1, \ldots, b^n \). By equation (3.1), the matrix representation of \( f \) with respect to this basis is \( B^T A B \).

Now suppose that the basis \( b^1, \ldots, b^n \) is orthonormal. We know from lecture 2 that in this case the matrix \( B \) is orthogonal, that is \( B^T B = I \).

Notice that if \( B \) and \( C \) are orthogonal \( NxN \) matrices, then \( BC \) is also orthogonal, since 
\[
(BC)^T (BC) = (C^T B^T) (BC) = C^T (B^T B) C = C^T C = I.
\]

The next thing to notice is that if \( A \) is a symmetric \( NxN \) matrix and \( B \) is an orthogonal \( NxN \) matrix, then \( \lambda \) is a characteristic value of \( A \) if and only if \( \lambda \) is a characteristic value of \( B^T A B \). This is so because \( \lambda \) is a characteristic value of \( A \) if and only if \( \det(A - \lambda I) = 0 \), which is true if and only if
\[
0 = (\det B^T) \det(A - \lambda I) \det B = \det(B^T A B - \lambda B^T B) = \det(B^T A B - \lambda I),
\]
which is true if and only if \( \lambda \) is a characteristic value of \( B^T A B \).

**Theorem 3.17:** If \( A \) is a symmetric \( NxN \) matrix, then there is an orthonormal basis for \( \mathbb{R}^n \) such that the matrix representation of the bilinear form \( f(x, y) = x^T Ay \) with respect to this basis is diagonal, where the diagonal entries are the characteristic values of \( A \). That is, there is an orthogonal matrix \( B \) such that 
\[
B^T A B = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \lambda_n
\end{pmatrix},
\]
where \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are characteristic values of \( A \) and include all the characteristic values of \( A \), perhaps with some of the \( \lambda_i \) being equal.

**Proof:** Let \( \lambda_1 \) be a characteristic value of \( A \) and let \( x_1 \) be a characteristic vector corresponding to \( \lambda_1 \). By lemma 3.16, \( \lambda_1 \) is a real number and hence we may assume that the \( N \)
components of $x^1_1$ are real. If necessary, replace $x^1_1$ by $x^1_1/||x^1_1||$, so that we may assume that $(x^1_1)^T x^1_1 = 1$. Use the Gram-Schmidt process to extend $x^1_1$ to an orthonormal basis, $x^1_1, x^1_2, \ldots, x^1_N$ of $R^N$. Let $B$ be the orthogonal matrix the columns of which are this basis and let $C = B^T A B$ and let $c_{1,11}^m$ be the $(m,n)$th entry of this matrix. Then

$$
 c_{1,11}^m = (x^1_1)^T A x^1_m = \lambda_1 (x^1_1)^T x^1_m = \lambda_1^m,
$$

If $m = 2, \ldots, N$, then $c_{1,1n}^m = (x^1_1)^T A x^1_n = \lambda_1 (x^1_1)^T x^1_n = 0$, since basis $x^1_1, x^1_2, \ldots, x^1_N$ is orthogonal. Since $C$ is symmetric, $c_{1,1n}^m = 0$, if $n = 2, \ldots, N$. That is, all the entries in the first column and row of $C$ are zero, except the upper left entry of the matrix, which equals $\lambda_1^1$. That is, the first row and column have been diagonalized.

We now proceed by induction on the number $k$ of rows and columns that have been diagonalized. Suppose by induction that we have defined orthogonal $N \times N$ matrices $B_1, \ldots, B_k$ such that $C = (B_1)^T (B_1) \ldots (B_k)^T A B \ldots B_k$ has the form

$$
 C_k = \begin{pmatrix}
 \lambda_1 & 0 & 0 & \cdots & 0 \\
 0 & \ddots & \ddots & \cdots & \vdots \\
 \vdots & \ddots & \ddots & \cdots & \vdots \\
 0 & 0 & \lambda_k & 0 & \cdots \\
 0 & 0 & c_{k,k+1,k+1} & c_{k,k+1,k+2} & \cdots \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 0 & 0 & c_{k,k+1,k+1} & \cdots & c_{k,k+1,N} \\
 0 & 0 & \cdots & \cdots & c_{k,k+1,N} \\
 \end{pmatrix}. 
$$

Let $C_{k+1}$ be the $(N-k) \times (N-k)$ matrix obtained by eliminating the first $k$ rows and columns of $C_k$. Let $\lambda_{k+1}$ be a characteristic value of $C_{k+1}$ and let $y_{k+1} \in R^{N-k}$ be a corresponding characteristic vector for $C_{k+1}$. We may assume that $(y_{k+1}^{k+1})^T y_{k+1}^{k+1} = 1$. Use the Gram-Schmidt process to extend $y_{k+1}^{k+1}$ to an orthonormal basis $y_{k+1}^{k+1}, \ldots, y_{nN}$ of $R^{N-k}$. For $n = k+1, \ldots, N$, let $x_n$ be the $N$-vector $(0, \ldots, 0, y_{n,k+1}^{k+1}, \ldots, y_{nN}^{nN})$ obtained from $y_n$ by adding zeros as the first $k$ components. Then $e_1, \ldots, e_n, x_{k+1}, \ldots, x_n$ is an orthonormal basis of $R^N$, where $e_n$ is the $n$th standard basis vector of $R^N$, for $n = 1, \ldots, k$. Let $B_{k+1}$ be the orthogonal $N \times N$ matrix whose columns are this basis. Then $C_{k+1} = (B_{k+1})^T C B_{k+1} = (B_{k+1})^T (B_{k+1}) \ldots (B_{k+1})^T A B \ldots B_{k+1}$ has the form
This completes the induction step.

Continue by induction until \( k = N \). We then have \( N \) orthogonal \( N \times N \) matrices \( B_1, \ldots, B_N \) such that the matrix \( C_k = (B_1)^T(B_2)^T \cdots (B_{k-1})^T B_k \) is diagonal. Since the product of orthogonal matrices is orthogonal, \( C_k = B^T_k A B_k \), where \( B = B_1 B_2 \cdots B_N \) is orthogonal. The matrices \( A \) and \( C_k \) have the same characteristic values, and the characteristic values of \( C_k \) are \( \lambda_1, \lambda_2, \ldots, \lambda_N \), the diagonal entries of \( C_k \).

**Proof of theorem 3.15:** I prove the theorem for the positive definite case, since the proofs for all four cases are nearly identical. By theorem 3.17, there is an orthogonal \( N \times N \) matrix \( B \) such that \( B^T A B \) is diagonal. \( B^T A B \) is positive definite if and only if \( A \) is positive definite. In order to see that this is so, notice that

\[
x^T (B^T A B) x = (B x)^T A (B x),
\]

so that if \( A \) is positive definite and \( x \) is any non-zero vector in \( \mathbb{R}^N \), then \( Bx \) is non-zero and so

\[
x^T (B^T A B) x = (B x)^T A (B x) > 0 \text{ and hence } B^T A B \text{ is positive definite}.
\]

Similarly

\[
x^T A x = x^T B B^T A B x,
\]

so that if \( B^T A B \) is positive definite and \( x \) is any non-zero vector in \( \mathbb{R}^N \), then \( B^T x \) is non-zero, so that

\[
x^T B (B^T A B) B^T x > 0 \text{ and hence } x^T Ax > 0 \text{ and so } A \text{ is positive definite}.
\]

If \( B^T A B \) is diagonal, then it is positive definite if and only if the diagonal elements are positive. To see that this is so, let \( \lambda_1, \lambda_2, \ldots, \lambda_N \) be the diagonal entries of \( B^T A B \). Then

\[
x^T B^T A B x = \sum_{n=1}^N \lambda_n x_n^2,
\]

which is positive if all \( x \neq 0 \) and \( \lambda_n > 0 \), for all \( n \).
1 show that if $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ and if $\theta$ is the angle between $x$ and $y$, then
\[
\cos \theta = \frac{x \cdot y}{\|x\| \|y\|},
\]
a fact listed in the section of lecture 2 on the inner product. First of all, I define the angle $\theta$ precisely. Let $b^1 = \frac{x}{\|x\|}$ be the vector of unit length in the direction of $x$ and let $b^2 = \frac{y - (y \cdot b^1) b^1}{\|y - (y \cdot b^1) b^1\|}$ be the vector of unit length orthogonal to $x$ and on the same side of $x$ as $y$ in the plane defined by $x$ and $y$. Notice that $(y \cdot b^1) b^1$ is the projection of $y$ onto the line through $x$ and that $b^1$ and $b^2$ form an orthonormal basis for the plane determined by $x$ and $y$, which is the same as the linear span of $x$ and $y$. By lemma 2.14, $y = (y \cdot b^1) b^1 + (y \cdot b^2) b^2$. Let $X = (\|x\|, 0, \ldots, 0) \in \mathbb{R}^N$ and $Y = (y \cdot b^1, y \cdot b^2, 0, \ldots, 0) \in \mathbb{R}^N$. If we draw $X$ and $Y$ in $\mathbb{R}^2$, as above, we see that
\[
\cos \theta = \frac{Y_1}{\|Y\|} = \frac{X \cdot Y}{\|X\| \|Y\|}.
\]
The question is, if $\theta$ is so defined, does $\cos \theta = \frac{x \cdot y}{\|x\| \|y\|}$? Using the Gram-Schmidt process,
\[
\|X\| \|Y\|
\]
we can extend the vectors $b^1$, $b^2$ to an orthonormal basis $b^1$, $b^2$, $\ldots$, $b^N$ for $\mathbb{R}^N$. Let $B$ be the $N \times N$ matrix the $n$th column of which is $b^n$, for all $n$. That is $B = (b^1 \ b^2 \ b^3 \ \ldots \ b^N)$. Since $b^1$, $b^2$, $\ldots$, $b^N$ is an orthonormal basis, $B^T B = I$, where $I$ is the $N \times N$ identity matrix. That is, $B$ is an orthogonal matrix. Then $x = BX$ and $y = BY$, so that by theorem 2.18, $X \cdot Y = x \cdot y$, $X \cdot x = x \cdot x$, and $Y \cdot Y = y \cdot y$. It follows that $\frac{X \cdot Y}{\|X\| \|Y\|} = \frac{x \cdot y}{\|x\| \|y\|}$, as was to be shown.