CHARACTERISTIC ROOTS AND VECTORS

1. Definition of characteristic roots and vectors

1.1. **Statement of the characteristic root problem.** Find values of a scalar λ for which there exist vectors x $\neq 0$ satisfying

$$Ax = \lambda x \tag{1}$$

where A is a given nth order matrix. The values of λ that solve the equation are called the characteristic roots or eigenvalues of the matrix A. To solve the problem rewrite the equation as

$$\begin{aligned} Ax &= \lambda x = \lambda I x \\ \Rightarrow (\lambda I - A)x &= 0 \qquad x \neq 0 \end{aligned} \tag{2}$$

For a given λ, any x which satisfies 1 will satisfy 2. This gives a set of n homogeneous equations in n unknowns. The set of x's for which the equation is true is called the null space of the matrix $(\lambda I - A)$. This equation can have a non-trivial solution iff the matrix $(\lambda I - A)$ is singular. This equation is called the characteristic or the determinantal equation of the matrix A. To see why the matrix must be singular consider a simple 2x2 case. First solve the system for x_1

$$\begin{aligned} Ax &= 0 \\ \Rightarrow a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_1 + a_{22}x_2 &= 0 \\ \Rightarrow x_1 = -\frac{a_{12}x_2}{a_{11}} \end{aligned} \tag{3}$$

Now substitute x_1 in the second equation

$$\begin{aligned} -a_{21}\frac{a_{12}x_2}{a_{11}} + a_{22}x_2 &= 0 \\ \Rightarrow x_2 \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right) &= 0 \\ \Rightarrow x_2 = 0 \; or \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right) &= 0 \\ If \; x_2 \neq 0 \; then \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right) &= 0 \\ \Rightarrow |A| &= 0 \end{aligned} \tag{4}$$

Date: September 16, 2004.

1.2. **Determinantal equation used in solving the characteristic root problem.** Now consider the singularity condition in more detail

$$
\begin{aligned}
(\lambda I - A)x &= 0 \\
\Rightarrow |\lambda I - A| &= 0 \\
\Rightarrow \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} &= 0
\end{aligned} \tag{5}
$$

This equation is a polynomial in λ since the formula for the determinant is a sum containing n! terms, each of which is a product of n elements, one element from each column of A. The fundamental polynomials are given as

$$
|\lambda I - A| = \lambda^n + b_{n-1}\lambda^{n-1} + b_{n-2}\lambda^{n-2} + \cdots + b_1\lambda + b_0 \tag{6}
$$

This is obvious since each row of $|\lambda I - A|$ contributes one and only one power of λ as the determinant is expanded. Only when the permutation is such that column included for each row is the same one will each term contain λ, giving λ^n. Other permutations will give lesser powers and b_0 comes from the product of the terms on the diagonal (not containing λ) of A with other members of the matrix. The fact that b_0 comes from all the terms not involving λ implies that it is equal to $|-A|$.

Consider a 2x2 example

$$
\begin{aligned}
|\lambda I - A| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{vmatrix} \\
&= (\lambda - a_{11})(\lambda - a_{22}) - (a_{12}\,a_{21}) \\
&= \lambda^2 - a_{11}\lambda - a_{22}\lambda + a_{11}\,a_{22} - a_{12}\,a_{21} \\
&= \lambda^2 + (-a_{11} - a_{22})\lambda + a_{11}\,a_{22} - a_{12}\,a_{21} \\
&= \lambda^2 - \lambda\,(a_{11} + a_{22}) + (a_{11}\,a_{22} - a_{12}\,a_{21}) \\
&= \lambda^2 + b_1\lambda + b_0 \\
b_0 &= |-A|
\end{aligned} \tag{7}
$$

Consider also a 3x3 example where we find the determinant using the expansion of the first row

$$
\begin{aligned}
|\lambda I - A| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \end{vmatrix} \\
&= (\lambda - a_{11}) \begin{vmatrix} \lambda - a_{22} & -a_{23} \\ -a_{32} & \lambda - a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} -a_{21} & -a_{23} \\ -a_{31} & \lambda - a_{33} \end{vmatrix} - a_{13} \begin{vmatrix} -a_{21} & \lambda - a_{22} \\ -a_{31} & -a_{32} \end{vmatrix}
\end{aligned} \tag{8}
$$

Now expand each of the three determinants in equation 8. We start with the first term

$$
\begin{aligned}
(\lambda - a_{11}) \begin{vmatrix} \lambda - a_{22} & -a_{23} \\ -a_{32} & \lambda - a_{33} \end{vmatrix} &= (\lambda - a_{11}) \left[\lambda^2 - \lambda a_{33} - \lambda a_{22} + a_{22} a_{33} - a_{23} a_{32} \right] \\
&= (\lambda - a_{11}) \left[\lambda^2 - \lambda (a_{33} + a_{22}) + a_{22} a_{33} - a_{23} a_{32}) \right] \\
&= \lambda^3 - \lambda^2 (a_{33} + a_{22}) + \lambda (a_{22} a_{33} - a_{23} a_{32}) - \lambda^2 a_{11} + \lambda a_{11} (a_{33} + a_{22}) - a_{11} \ (a_{22} a_{33} - a_{23} a_{32}) \\
&= \lambda^3 - \lambda^2 (a_{11} + a_{22} + a_{33}) + \lambda (a_{11} a_{33} + a_{11} a_{22} + a_{22} a_{33} - a_{23} a_{32}) - a_{11} a_{22} a_{33} + a_{11} a_{23} a_{32}
\end{aligned} \tag{9}
$$

Now the second term

$$
\begin{aligned}
a_{12} \begin{vmatrix} -a_{21} & -a_{23} \\ -a_{31} & \lambda - a_{33} \end{vmatrix} &= a_{12} \left[-\lambda a_{21} + a_{21} a_{33} - a_{23} a_{31} \right] \\
&= - \lambda a_{12} a_{21} + a_{12} a_{21} a_{33} - a_{12} a_{23} a_{31}
\end{aligned} \tag{10}
$$

Now the third term

$$
\begin{aligned}
- a_{13} \begin{vmatrix} -a_{21} & \lambda - a_{22} \\ -a_{31} & -a_{32} \end{vmatrix} &= - a_{13} \left[a_{21} a_{32} + \lambda a_{31} - a_{22} a_{31} \right] \\
&= - a_{13} a_{21} a_{32} - \lambda a_{13} a_{31} + a_{13} a_{22} a_{31}
\end{aligned} \tag{11}
$$

Now combine the three expressions to obtain

$$
\begin{aligned}
|\lambda I - A| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \end{vmatrix} \\
&= \lambda^3 - \lambda^2 (a_{11} + a_{22} + a_{33}) + \lambda (a_{11} a_{33} + a_{11} a_{22} + a_{22} a_{33} - a_{23} a_{32}) - a_{11} a_{22} a_{33} + a_{11} a_{23} a_{32} \\
&\quad - \lambda a_{12} a_{21} + a_{12} a_{21} a_{33} - a_{12} a_{23} a_{31} \\
&\quad - a_{13} a_{21} a_{32} - \lambda a_{13} a_{31} + a_{13} a_{22} a_{31} \\
&= \lambda^3 - \lambda^2 (a_{11} + a_{22} + a_{33}) + \lambda (a_{11} a_{33} + a_{11} a_{22} + a_{22} a_{33} - a_{23} a_{32} - a_{12} a_{21} - a_{13} a_{31}) \\
&\quad - a_{11} a_{22} a_{33} + a_{11} a_{23} a_{32} + a_{12} a_{21} a_{33} - a_{12} a_{23} a_{31} - a_{13} a_{21} a_{32} + a_{13} a_{22} a_{31}
\end{aligned} \tag{12}
$$

The first term will be λ^3, the others will give polynomials in λ^2, and λ. Note that the constant term is the negative of the determinant of A.

1.3. **Fundamental theorem of algebra.**

1.3.1. *Statement of Fundamental Theorem Theorem of Algebra.*

Theorem 1. *Any polynomial p(x) of degree at least 1, with complex coefficients has at least one zero z (i.e., z is a root of the equation* **p(x) = 0** *among the complex numbers. Further, if z is a zero of p(x), then x - z divides p(x) and p(x) = (x-z)q(x), where q(x) is a polynomial with complex coefficients, whose degree is 1 smaller than p.*

What this says is that we can write the polynomial as a product of (x-z) and a term which is a polynomial of one less degree than p(x). For example if p(x) is a polynomial of degree 4 then we can write it as

$$
p(x) = (x - x_1) * (\text{polynomial with power no greater than 3}) \tag{13}
$$

where x_1 is a root of the equation. Given that q(x) is a polynomial and if it is of degree greater than one, then we can write it in a similar fashion as

$$q(x) \; = \; (x - x_2) * \text{(polynomial with power no greater than 2)} \tag{14}$$

where x_2 is a root of the equation q(x) = 0. This then implies that p(x) can be written as

$$p(x) \; = \; (x - x_1)\,(x - x_2) * \text{(polynomial with power no greater than 2)} \tag{15}$$

or continuing

$$p(x) \; = \; (x - x_1)\,(x - x_2)\,(x - x_3) * \text{(polynomial with power no greater than 1)} \tag{16}$$

$$p(x) \; = \; (x - x_1)\,(x - x_2)\,(x - x_3)\,(x - x_4) * \text{(term not containing x)} \tag{17}$$

If we set this equation equal to zero, it implies that

$$p(x) \; = \; (x - x_1)\,(x - x_2)\,(x - x_3)\,(x - x_4) \; = \; 0 \tag{18}$$

1.3.2. *Example of Fundamental Theorem Theorem of Algebra.* Consider the equation

$$p(t) \; = \; t^3 \; - \; 6t^2 \; + \; 11t \; - \; 6 \tag{19}$$

This has roots t^1 = 1, t^2 = 2, t^3 =3. Consider then that we can write the equation as (t-1)q(t) as follows

$$\begin{aligned} p(t) \; &= (t-1)(t^2 \; - \; 5t \; + \; 6) \\ &= t^3 \; - \; 5t^2 \; + \; 6t \; - \; t^2 \; + \; 5t \; - \; 6 \\ &= t^3 \; - \; 6t^2 \; + \; 11t \; - \; 6 \end{aligned} \tag{20}$$

Now carry this one step further as

$$\begin{aligned} q(t) \; &= t^2 \; - \; 5t \; + \; 6 \\ &= (t-2)s(t) \; = \; (t-2)(t-3) \\ \Rightarrow \; p(t) \; &= (t \; - \; t_1)(t \; - \; t_2)(t \; - \; t_3) \end{aligned} \tag{21}$$

1.3.3. *Theorem that Follows from the Fundamental Theorem Theorem of Algebra.*

Theorem 2. *A polynomial of degree $n \; \geq \; 1$ with complex coefficients has, counting multiplicities, exactly n zeroes among the complex numbers.*

The multiplicity of a root z of p(x) = 0 is the largest integer k for which (x-z)^k divides p(x), that is, the number of times z occurs as a root of p(x) = 0. If z has a multiplicity of 2, then it is counted twice (2 times) toward the number n of roots of p(x) = 0.

1.3.4. *First example of theorem 2.* Consider the polynomial

$$p(t) \;=\; t^3 \;-\; 6t^2 \;+\; 11t \;-\; 6 \tag{22}$$

If we divide this by (t-1) we obtain

$$\begin{aligned} \frac{p(t)}{(t-1)} \;&=\; \frac{(t^3 \;-\; 6t^2 \;+\; 11t \;-\; 6)}{(t-1)} \\ &= t^2 \;-\; 5t \;+\; 6 \\ &=(t-2)(t-3) \end{aligned} \tag{23}$$

which is not divisible by (t-1), so the multiplicity of the root 1 is 1.

1.3.5. *Second example of theorem 2.* Consider the polynomial

$$p(t) \;=\; t^3 \;-\; 5t^2 \;+\; 8t \;-\; 4 \tag{24}$$

One root of this is 2 because

$$\begin{aligned} p(t) \;&= t^3 \;-\; 5t^2 \;+\; 8t \;-\; 4 \\ &= 2^3 \;-\; 5*2^2 \;+\; 8*2 \;-\; 4 \\ &= 8 \;-\; 20 \;+\; 16 \;-\; 4 \;=\; 0 \end{aligned} \tag{25}$$

If we divide the polynomial by (t-2) we obtain

$$\begin{aligned} \frac{p(t)}{(t-2)} \;&=\; \frac{t^3 \;-\; 5t^2 \;+\; 8t \;-\; 4}{(t-2)} \\ &= t^2 \;-\; 3t \;+\; 2 \\ &= (t-2)(t-1) \end{aligned} \tag{26}$$

which is divisible by (t-2) so the multiplicity of the root 2 is 2.

1.3.6. *factoring nth degree polynomials.* An nth degree polynomial $f(\lambda)$ can be written in factored form using the solutions of the equation $f(\lambda) = 0$ as

$$\begin{aligned} f(\,\lambda\,) \;&= (\lambda \;-\; \lambda_1)(\lambda \;-\; \lambda_2)\;\cdots\;(\lambda \;-\; \lambda_n) \;=\; 0 \\ &= \prod_{i=1}^{n}\,(\,\lambda - \lambda i\,) \;=\; 0 \end{aligned} \tag{27}$$

If we multiply this out for the special case when the polynomial is of degree 4 we obtain

$$\begin{aligned} f(\,\lambda\,) \;&=\; (\lambda \;-\; \lambda_1)(\lambda \;-\; \lambda_2)(\lambda \;-\; \lambda_3)(\lambda \;-\; \lambda_4) \;=\; 0 \\ \Rightarrow\; (\lambda \;-\; \lambda_1) &* \text{term containing no power of } \lambda \text{ larger than 3} \;=\; 0 \\ \Rightarrow\; \lambda\;[(\lambda \;-\; \lambda_2)(\lambda \;-\; \lambda_3)(\lambda \;-\; \lambda_4] \;-\; \lambda_1\;[(\lambda \;-\; \lambda_2)&(\lambda \;-\; \lambda_3)(\lambda \;-\; \lambda_4] \;=\; 0 \end{aligned} \tag{28}$$

Examining equation 27 or equation 28, we can see that we will eventually end up with a sum of n terms, the first being λ^4, the second being λ^3 multiplied by a term involving $\lambda_1, \lambda_2, \lambda_3$, and λ_4, the third being λ^2 multiplied by a term involving $\lambda_1, \lambda_2, \lambda_3$, and λ_4, and so on until the last term which will be $\prod_{i=1}^{4} \lambda i$. Completing the computations for the 4th degree polynomial will yield

$$
\begin{aligned}
f(\lambda) &= \lambda^4 + \lambda^3(-\lambda_1 - \lambda_2 - \lambda_3 - \lambda_4) \\
&+ \lambda^2(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 + \lambda_1\lambda_4\lambda_2\lambda_4 + \lambda_3\lambda_4) \\
&+ \lambda(-\lambda_1\lambda_2\lambda_3 - \lambda_1\lambda_2\lambda_4 - \lambda_1\lambda_3\lambda_4 - \lambda_2\lambda_3\lambda_4) + (\lambda_1\lambda_2\lambda_3\lambda_4) = 0 \\
&= \lambda^4 - \lambda^3(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) \\
&+ \lambda^2(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 + \lambda_1\lambda_4 + \lambda_2\lambda_4 + \lambda_3\lambda_4) \\
&- \lambda(\lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_3\lambda_4 + \lambda_2\lambda_3\lambda_4) + (\lambda_1\lambda_2\lambda_3\lambda_4) = 0
\end{aligned}
\tag{29}
$$

We can write this in an alternative way as follows

$$
\begin{aligned}
f(\lambda) &= \lambda^4 - \lambda^3\, \Sigma_{i=1}^4 \lambda_i \\
&+ \lambda^2\, \Sigma_{i\neq j} \lambda_i\lambda_j \\
&- \lambda^1\, \Sigma_{i\neq j\neq k} \lambda_i\lambda_j\lambda_k \\
&+ (-1)^4\, \Pi_{i=1}^4 \lambda_i = 0
\end{aligned}
\tag{30}
$$

Similarly for polynomials of degree 2, 3 and 5, we find

$$
\begin{aligned}
f(\lambda) &= \lambda^2 - \lambda\, \Sigma_{i=1}^2 \lambda_i \\
&+ (-1)^2\, \Pi_{i=1}^2 \lambda_i = 0
\end{aligned}
\tag{31a}
$$

$$
\begin{aligned}
f(\lambda) &= \lambda^3 - \lambda^2\, \Sigma_{i=1}^3 \lambda_i \\
&+ \lambda\, \Sigma_{i\neq j} \lambda_i\lambda_j \\
&+ (-1)^3\, \Pi_{i=1}^3 \lambda_i = 0
\end{aligned}
\tag{31b}
$$

$$
\begin{aligned}
f(\lambda) &= \lambda^5 - \lambda^4\, \Sigma_{i=1}^5 \lambda_i \\
&+ \lambda^3\, \Sigma_{i\neq j} \lambda_i\lambda_j \\
&- \lambda^2\, \Sigma_{i\neq j\neq k} \lambda_i\lambda_j\lambda_k \\
&+ \lambda^1\, \Sigma_{i\neq j\neq k\neq \ell} \lambda_i\lambda_j\lambda_k\lambda_\ell \\
&+ (-1)^4\, \Pi_{i=1}^4 \lambda_i = 0
\end{aligned}
\tag{31c}
$$

Now consider the general case.

$$
f(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\, \ldots\, (\lambda - \lambda_n) = 0 \tag{32}
$$

Each product contributing to the power λ^k in the expansion is the result of k times choosing λ and $n-k$ times choosing one of the λ_i. For example, with the first term of a third degree polynomial, we choose λ three times and do not choose any of the λ_i. For the second term we choose λ twice and each of the λ_i once. When we choose λ once we choose each of the λ_i twice. The term not involving λ implies that we choose all of the λ_i. This is of course $(\lambda_1\, \lambda_2, \lambda_3)$. Choosing λ k times and λ_i $(n-k)$ times is a problem in combinatorics. The answer is given by the binomomial formula, specifically there are $\binom{n}{n-k}$ products with the power λ^k. In a cubic, when k = 3, there are no terms involving λ_i, when k = 2 there will three terms involving the individual λ_i, when k = 1 there will three terms involving the product of two of the λ_i, and when k = 0 there will one term involving the product of three of the λ_i as can be seen in equation 33.

$$\begin{aligned} f(\lambda) &= (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \\ &= (\lambda^2 - \lambda\lambda_1 - \lambda\lambda_2 + \lambda_1\lambda_2)(\lambda - \lambda_3) \\ &= \lambda^3 - \lambda^2\lambda_1 - \lambda^2\lambda_2 + \lambda\lambda_1\lambda_2 - \lambda^2\lambda_3 + \lambda\lambda_1\lambda_3 + \lambda\lambda_2\lambda_3 - \lambda_1\lambda_2\lambda_3 \\ &= \lambda^3 - \lambda^2(\lambda_1 + \lambda_2 + \lambda_3) + \lambda(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3) - \lambda_1\lambda_2\lambda_3 \end{aligned} \tag{33}$$

The coefficient of λ^k will be the sum of the appropriate products, multiplied by -1 if an odd number of λ_i have been chosen. The general case follows

$$\begin{aligned} f(\lambda) &= (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n) \\ &= \lambda^n - \lambda^{n-1}\Sigma_{i=1}^n \lambda_i \\ &+ \lambda^{n-2}\Sigma_{i\neq j}\lambda_i\lambda_j \\ &- \lambda^{n-3}\Sigma_{i\neq j\neq k}\lambda_i\lambda_j\lambda_k \\ &+ \lambda^{n-4}\Sigma_{i\neq j\neq k\neq \ell}\lambda_i\lambda_j\lambda_k\lambda_\ell \\ &+ \ldots \\ &+ (-1)^n\Pi_{i=1}^n\lambda_i = 0 \end{aligned} \tag{34}$$

The term not containing λ is always $(-1)^n$ times the product of the roots $(\lambda_1, \lambda_2, \lambda_3, \ldots)$ of the polynomial and term containing λ^{n-1} is always $(-\sum_{i=1}^n \lambda_i)$.

1.3.7. *Implications of factoring result for the fundamental determinantal equation 6.* The fundamental equations for computing characteristic roots are

$$(\lambda I - A)x = 0 \tag{35a}$$

$$|\lambda I - A| = 0 \tag{35b}$$

$$|\lambda I - A| = \lambda^n + b_{n-1}\lambda^{n-1} + b_{n-2}\lambda^{n-2} + \ldots + b_1\lambda + b_0 = 0 \tag{35c}$$

Equation 35c is just a polynomial in λ. If we solve the equation, we will obtain n roots (some perhaps repeated). Once we find these roots, we can also write equation 35c in factored form using equation 34. It is useful to write the two forms together.

$$|\lambda I - A| = \lambda^n + b_{n-1}\lambda^{n-1} + b_{n-2}\lambda^{n-2} + \ldots + b_1\lambda + b_0 = 0 \tag{36a}$$

$$|\lambda I - A| = \lambda^n - \lambda^{n-1}\Sigma_{i=1}^n\lambda_i + \lambda^{n-2}\Sigma_{i\neq j}\lambda_i\lambda_j + \ldots + (-1)^n\Pi_{i=1}^n\lambda_i = 0 \tag{36b}$$

We can then find the coefficients of the various powers of λ by comparing the two equations. For example, $b_{n-1} = -\Sigma_{i=1}^n\lambda_i$ and $b_0 = (-1)^n\Pi_{i=1}^n\lambda_i$.

1.3.8. *Implications of theorem 1 and theorem 2.* The n roots of a polynomial equation need not all be different, but if a root is counted the number of times equal to its multiplicity, there are n roots of the equation. Thus there are n roots of the characteristic equation since it is an nth degree polynomial. These roots may all be different or some may be the same. The theorem also implies that there can not be more than n distinct values of λ for which $|\lambda I - A| = 0$. For values of λ different than the roots of $f(\lambda)$, solutions of the characteristic equation require x = 0. If we set λ equal to one of the roots of the equation, say λ_i, then $|\lambda_i I - A| = 0$.

For example, consider a matrix A as follows

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \tag{37}$$

This implies that

$$|\,\lambda I \,-\, A\,| = \begin{vmatrix} \lambda - 4 & -2 \\ -2 & \lambda - 4 \end{vmatrix} \tag{38}$$

Taking the determinant we obtain

$$\begin{aligned} |\,\lambda I \,-\, A\,| &= (\lambda - 4)(\lambda - 4) - 4 \\ &= \lambda^2 \,-\, 8\,\lambda \,+\, 12 \,=\, 0 \\ &\Rightarrow (\lambda \,-\, 6)\,(\lambda \,-\, 2) \,=\, 0 \\ &\Rightarrow \lambda_1 \,=\, 6,\; \lambda_2 \,=\, 2 \end{aligned} \tag{39}$$

Now write out equation 39 with λ_1 = 2 as follows

$$\begin{aligned} A &= \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} \\ |\,2\,I \,-\, A\,| &= \begin{vmatrix} 2 - 4 & -2 \\ -2 & 2 - 4 \end{vmatrix} \\ &= \begin{vmatrix} -2 & -2 \\ -2 & -2 \end{vmatrix} \\ &= 0 \end{aligned} \tag{40}$$

1.4. **A numerical example of computing characteristic roots (eigenvalues).** Consider a simple example matrix

$$
\begin{aligned}
A &= \begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix} \\
|\,\lambda I \,-\, A\,| &= \begin{vmatrix} \lambda - 4 & -2 \\ -3 & \lambda - 5 \end{vmatrix} \\
&= (\lambda - 4)(\lambda - 5) - 6 \\
&= \lambda^2 \,-\, 4\lambda \,-\, 5\lambda \,+\, 20 \,-\, 6 \\
&= \lambda^2 \,-\, 9\lambda \,+\, 14 \,= 0 \\
&\Rightarrow (\lambda \,-\, 7)\,(\lambda \,-\, 2) \,=\, 0 \\
&\Rightarrow \lambda_1 \,=\, 7,\; \lambda_2 \,= 2
\end{aligned}
\tag{41}
$$

1.5. **Characteristic vectors.** Now return to the general problem. Values of λ which solve the determinantal equation are called the characteristic roots or eigenvalues of the matrix A. Once λ is known, we may be interested in vectors x which satisfy the characteristic equation. In examining the general problem in equation 1, it is clear that if x satisfies the equation for a given λ, so will cx, where c is a scalar. Thus x is not determined. To remove this indeterminacy, a common practice is to normalize the vector x such that x′x = 1. Thus the solution actually consists of λ and n-1 elements of x. If there are multiple values of λ, then there will be an x vector associated with each.

1.6. **Computing characteristic vectors (eigenvectors).** To find a characteristic vector, substitute a characteristic root in the determinantal equation and then solve for x. The solution will not be determinate, but will require the normalization. Consider the example in equation 41 with λ_1 = 2. Making the substitution and writing out the determinantal equation will yield.

$$
\begin{aligned}
A &= \begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix} \\
[\lambda I \,-\, A]x &= \begin{pmatrix} 2 - 4 & -2 \\ -3 & 2 - 5 \end{pmatrix} x \\
&= \begin{pmatrix} -2 & -2 \\ -3 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,=\, 0
\end{aligned}
\tag{42}
$$

Now solve the system of eqautions for x_1

$$
\begin{aligned}
-2x_1 \,-\, 2x_2 &= 0 \\
-3x_1 \,-\, 3x_2 &= 0 \\
\Rightarrow\; x_1 \,=\, -\,x_2 \; and \; x_1^2 \,+\, x_2^2 \,&=\, 1
\end{aligned}
\tag{43}
$$

Substitute for x_1 in equation 42

$$
\begin{aligned}
(-x_2)^2 \,+\, x_2^2 &= 1 \\
\Rightarrow\; 2x_2^2 &= 1 \\
\Rightarrow\; x_2^2 &= \frac{1}{2} \\
\Rightarrow\; x_2 &= \frac{1}{\sqrt{2}},\; x_1 \,=\, -\frac{1}{\sqrt{2}}
\end{aligned}
\tag{44}
$$

Substitute the characteristic vector from equation 44 in the fundamental relationship given in equation 1

$$
\begin{aligned}
Ax &= \begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 2 \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{-2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{-2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix}
\end{aligned} \tag{45}
$$

Similarly for $\lambda_2 = 7$

$$
\begin{aligned}
A &= \begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix} \\
[\lambda I - A]x &= \begin{pmatrix} 7-4 & -2 \\ -3 & 7-5 \end{pmatrix} x \\
&= \begin{pmatrix} 3 & -2 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \\
\Rightarrow 3x_1 - 2x_2 &= 0 \\
-3x_1 + 2x_2 &= 0 \\
\Rightarrow x_1 &= \frac{2}{3} x_2 \; and \; x_1^2 + x_2^2 = 1
\end{aligned} \tag{46}
$$

Substituting we obtain

$$
\begin{aligned}
\frac{4}{9} x_2^2 + x_2^2 &= 1 \\
\Rightarrow \frac{13}{9} x_2^2 &= 1 \\
\Rightarrow x_2^2 &= \frac{9}{13} \\
\Rightarrow x_2 &= \frac{3}{\sqrt{13}}, \; x_1 = \frac{2}{\sqrt{13}} \\
Ax &= \begin{pmatrix} 4 & 2 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{pmatrix} = \begin{pmatrix} \frac{14}{\sqrt{13}} \\ \frac{21}{\sqrt{13}} \end{pmatrix} \\
\lambda x &= 7x = \begin{pmatrix} \frac{14}{\sqrt{13}} \\ \frac{21}{\sqrt{13}} \end{pmatrix}
\end{aligned} \tag{47}
$$

2. Similar matrices

2.1. **Definition of Similar Matrices.**

Theorem 3. *Let A be a square matrix of order n and let Q be an invertible matrix of order n. Then*

$$B \;=\; Q^{-1}A\,Q \tag{48}$$

has the same characteristic roots as A, and if x is a characteristic vector of A, then $Q^{-1}x$ *is a characteristic vector of B. The matrices A and B are said to be similar.*

Proof. Let (λ, x) be a characteristic root and vector pair for the matrix A. Then

$$\begin{aligned} Ax &= \lambda x \\ Q^{-1}Ax &= \lambda\, Q^{-1}x \\ Also\; Q^{-1}A &= Q^{-1}AQQ^{-1} \\ So\; [Q^{-1}AQQ^{-1}]x &= \lambda\, [Q^{-1}x] \\ \Rightarrow\; Q^{-1}AQ\, [Q^{-1}x] &= \lambda\, [Q^{-1}x] \end{aligned} \tag{49}$$

□

This then implies that λ is a characteristic root of $Q^{-1}AQ$, with characteristic vector $Q^{-1}x$.

2.2. **Matrices similar to diagonal matrices.**

2.2.1. *Diagonizable matrix.* A matrix A is said to be diagonalizable if it is similar to a diagonal matrix. This means that there exists a matrix P such that $P^{-1}AP = D$, where D is a diagonal matrix.

2.2.2. *Theorem on diagonizable matrices.*

Theorem 4. *The nxn matrix A is diagonalizable iff there is a set of n linearly independent vectors, each of which is a characteristic vector of A.*

Proof. First show that if we have n linearly independent characteristic vectors, A is diagonalizable. Take the n linearly independent characteristic vectors of A, denoted x^1, x^2, ... x^n, with their characteristic roots, λ_1, ... λ_n and form a nonsingular matrix P with them as columns. So we have

$$P \;= [x^1\; x^2\; \cdots\; x^n] \;=\; \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ x_3^1 & x_3^2 & \cdots & x_3^n \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^n \end{bmatrix} \tag{50}$$

Let Λ be a diagonal matrix with the characteristic roots of A on the diagonal. Now calculate $P^{-1}AP$ as follows

$$
\begin{aligned}
P^{-1}AP &= P^{-1}[Ax^1\ Ax^2\ \cdots\ Ax^n] \\
&= P^{-1}[\lambda_1 x^1\ \lambda_2 x^2\ \cdots\ \lambda_n x^n] \\
&= P^{-1}[x^1\ x^2\ \cdots\ x^n] \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\
&= P^{-1}[x^1\ x^2\ \cdots\ x^n]\,\Lambda \\
&= P^{-1}P\Lambda \;=\; \Lambda
\end{aligned} \tag{51}
$$

Now suppose that A is diagonalizable so that $P^{-1}AP = D$. Then multiplying both sides by P we obtain AP = PD. This means that A times the ith column of P is the ith diagonal entry of D times the ith column of P. This means that the ith column of P is a characteristic vector of A associated with the ith diagonal entry of D. Thus D is equal to Λ. Since we assume that P is non-singular, the columns (characteristic vectors) are independent.

□

3. Independence of characteristic vectors

Theorem 5. *Suppose λ_1, ... ,λ_k are characteristic roots of an nxn matrix A, no two of which are the same. Let x^i be the characteristic vector associated with λ_i. Then the set [x^1, x^2, ... , x^k] is linearly independent.*

Proof. Suppose x^1, ..., x^k are linearly dependent. Then there exists a nontrivial linear combination such that

$$
a_1 \begin{pmatrix} x_1^1 \\ x_2^1 \\ \vdots \\ x_n^1 \end{pmatrix} + a_2 \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{pmatrix} + a_3 \begin{pmatrix} x_1^3 \\ x_2^3 \\ \vdots \\ x_n^3 \end{pmatrix} + \cdots + a_k \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_n^k \end{pmatrix} = 0 \tag{52}
$$

As an example the sum of the first two vectors might be zero. In this case a_1 = 1, a_2 = 1, and a_3, ... , a_k = 0. Consider the linear combination of the x's that has the fewest non-zero coefficients. Renumber the vectors such that these are the first r. Call the coefficients α. That is

$$
\alpha_1 x^1 + \alpha_2 x^2 + \ldots + \alpha_r x^r = 0\ for \quad r \leq k, \text{ and } \alpha_{r+1}, \ldots, \alpha_k = 0 \tag{53}
$$

Notice that $r > 1$, because all $x^i \neq 0$. Now multiply the matrix equation in 53 by A

$$
\begin{aligned}
A(\alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_r x^r) &= \alpha_1 A x^1 + \alpha_2 A x^2 + \cdots + \alpha_r A x^r \\
&= \alpha_1 \lambda_1 x^1 + \alpha_2 \lambda_2 x^2 + \cdots + \alpha_r \lambda_r x^r = 0
\end{aligned} \tag{54}
$$

Now multiply the result in 53 by λ_r and subtract from the last expression in 54. First multiply 53 by λ_r.

$$
\begin{aligned}
\lambda_r \left[\alpha_1 x^1 + \alpha_2 x^2 + \ldots + \alpha_r x^r \right] &= \lambda_r \alpha_1 x^1 + \lambda_r \alpha_2 x^2 + \ldots + \lambda_r \alpha_r x^r \\
&= \alpha_1 \lambda_r x^1 + \alpha_2 \lambda_r x^2 + \ldots + \alpha_r \lambda_r x^r = 0
\end{aligned} \tag{55}
$$

Now subtract

$$
\begin{array}{rrrrcl}
\alpha_1\lambda_1 x^1 & +\,\alpha_2\lambda_2 x^2\ +\ \cdots & +\,\alpha_{r-1}\lambda_{r-1}x^{r-1} & +\,\alpha_r\lambda_r x^r & = & 0 \\
\alpha_1\lambda_r x^1 & +\,\alpha_2\lambda_r x^2\ +\ \cdots & +\,\alpha_{r-1}\lambda_r x^{r-1} & +\alpha_r\lambda_r x^r & = & 0 \\
-----& ----- & ------- & ----- & & -- \\
\alpha_1(\lambda_1 - \lambda_r)x^1 & +\,\alpha_2(\lambda_2 - \lambda_r)x^r\ +\ \cdots & +\,\alpha_{r-1}(\lambda_{r-1} - \lambda_r)x^{r-1} & +\,\alpha_r(\lambda_r - \lambda_r)x^r & = & 0 \\
\Rightarrow\ \alpha_1(\lambda_1 - \lambda_r)x^1 & +\,\alpha_2(\lambda_2 - \lambda_r)x^2\ +\ \cdots & +\,\alpha_{r-1}(\lambda_{r-1} - \lambda_r)x^{r-1} & & = & 0
\end{array} \tag{56}
$$

This dependence relation has fewer coefficients than 53, thus contradicting the assumption that 53 contains the minimal number. Thus the vectors x^1, ... , x^k must be independent.

□

4. Distinct characteristic roots and diagonalizability

Theorem 6. *If an nxn matrix A has n distinct characteristic roots, then it is diagonalizable.*

Proof. Let the distinct characteristic roots of A be given by λ_1, ... , λ_n. Since they are all different, x^1, ..., $x^n$is a linearly independent set by the last theorem. Then A is diagonalizable by the Theorem 4.

□

5. Determinants, traces, and characteristic roots

5.1. Theorem on determinants.

Theorem 7. *Let A be a square matrix of order n. Let λ_1, ..., λ^n be its characteristic roots. Then*

$$
|\, A \,| \;=\; \Pi_{i=1}^{n}\, \lambda_i \tag{57}
$$

Proof. Consider the n roots of the equation defining the characteristic roots. Consider the general equation for the roots of an equation (as in equation 6, the example in equation 12 or equation 34) and the specific case here.

$$
\begin{aligned}
|\,\lambda I \,-\, A\,| \;=\; (\lambda - \lambda_1)(\lambda - \lambda_2)\,\ldots\,(\lambda - \lambda_n) \;=0 \\
\Rightarrow\ \lambda^n \,+\, \lambda^{n-1}\,(\text{terms involving } \lambda_i\ ) \,+\, \cdots \,+\, \lambda\,(\text{terms involving } \lambda_i\ ) \,+\, \prod_{i=1}^{n}\,(-\,\lambda_i\,) \;=0 \\
\Rightarrow\ \lambda^n - \lambda^{n-1}\,\Sigma_{i=1}^{n-1}\,\lambda_i \;+\; \lambda^{n-2}\,\Sigma_{i\neq j}\lambda_i\lambda_j \;-\; \lambda^{n-3}\,\Sigma_{i\neq j\neq k}\lambda_i\lambda_j\lambda_k \\
+\,\lambda^{n-4}\,\Sigma_{i\neq j\neq k\neq \ell}\lambda_i\lambda_j\lambda_k\lambda_\ell \,-\, \cdots \,+\, (-1)^n\,\Pi_{i=1}^{n}\,\lambda_i \;=0
\end{aligned} \tag{58}
$$

The first term is λ^n and the last term contains only the λ_i. This last term is given by

$$
\Pi_{i=1}^{n}\,(-\lambda_i) \;=\; (-1)^n\,\Pi_{i=1}^{n}\lambda_i \tag{59}
$$

Now compare equation 6 and equation 58

$$
(6)\ |\,\lambda I - A\,| \;=\lambda^n \quad +\,b_{n-1}\lambda^{n-1} \quad +\,b_{n-2}\lambda^{n-2} \quad +\ldots \quad +\,b_0 \quad =0 \tag{60a}
$$

$$
(58)\ |\,\lambda I - A\,| \;=\lambda^n \quad -\lambda^{n-1}\,\Sigma_{i=1}^{n-1}\,\lambda_i \quad +\lambda^{n-2}\,\Sigma_{i\neq j}\lambda_i\lambda_j \quad -\cdots \quad +\,(-1)^n\,\Pi_{i=1}^{n}\,\lambda_i \quad =0 \tag{60b}
$$

Notice that the last expression in 58 must be equivalent to the b_0 in equation 6 because it is the only term not containing λ. Specifically,

$$
b_0 \;=\; (-1)^n\,\Pi_{i=1}^{n}\,\lambda_i \tag{61}
$$

Now set $\lambda = 0$ in 60a. This will give

$$|\,\lambda I - A\,| \;=\; \lambda^n + b_{n-1}\lambda^{n-1} + b_{n-2}\lambda^{n-2} + \ldots + b_0 = 0$$
$$\Rightarrow \;|\,-A\,| = b_0 \tag{62}$$
$$\Rightarrow \;(-1)^n\,|\,A\,| = b_0$$

Therefore

$$b_0 \;= (-1)^n\,\Pi_{i=1}^n\,\lambda_i \;=\; (-1)^n\,|\,A\,| \;=\; b_0$$
$$\rightarrow\; (-1)^n\,\Pi_{i=1}^n\,\lambda_i \;= (-1)^n\,|\,A\,| \tag{63}$$
$$\rightarrow\; \Pi_{i=1}^n\,\lambda_i \;= |\,A\,|$$

□

5.2. Theorem on traces.

Theorem 8. *Let A be a square matrix of order n. Let* λ_1*, ...,* λ^n *be its characteristic roots. Then*

$$tr\ A \;=\; \Sigma_{i=1}^n\,\lambda_i \tag{64}$$

To help visualize the proof, consider the following case where n = 4.

$$|\,\lambda I \;-\; A\,| \;=\; \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{21} & \lambda - a_{22} & -a_{23} & -a_{24} \\ -a_{31} & -a_{32} & \lambda - a_{33} & -a_{34} \\ -a_{41} & -a_{42} & -a_{43} & \lambda - a_{44} \end{vmatrix}$$

$$= (\lambda - a_{11}) \begin{vmatrix} \lambda - a_{22} & -a_{23} & -a_{24} \\ -a_{32} & \lambda - a_{33} & -a_{34} \\ -a_{42} & -a_{43} & \lambda - a_{44} \end{vmatrix} \;+\; a_{12} \begin{vmatrix} -a_{21} & -a_{23} & -a_{24} \\ -a_{31} & \lambda - a_{33} & -a_{34} \\ -a_{41} & -a_{43} & \lambda - a_{44} \end{vmatrix} \tag{65}$$

$$+ (-a_{13}) \begin{vmatrix} -a_{21} & \lambda - a_{22} & -a_{24} \\ -a_{31} & -a_{32} & -a_{34} \\ -a_{41} & -a_{42} & \lambda - a_{44} \end{vmatrix} \;+\; a_{14} \begin{vmatrix} -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \\ -a_{41} & -a_{42} & -a_{43} \end{vmatrix}$$

Also consider expanding the determinant

$$\begin{vmatrix} \lambda - a_{22} & -a_{23} & -a_{24} \\ -a_{32} & \lambda - a_{33} & -a_{34} \\ -a_{42} & -a_{43} & \lambda - a_{44} \end{vmatrix} \tag{66}$$

by its first row.

$$(\lambda - a_{22}) \begin{vmatrix} \lambda - a_{33} & -a_{34} \\ -a_{43} & \lambda - a_{44} \end{vmatrix} \;+\; a_{23} \begin{vmatrix} -a_{32} & -a_{34} \\ -a_{42} & \lambda - a_{44} \end{vmatrix} \;+\; (-a_{24}) \begin{vmatrix} -a_{32} & \lambda - a_{33} \\ -a_{42} & -a_{43} \end{vmatrix} \tag{67}$$

Now proceed with the proof.

Proof. Expanding the determinantal equation (5) by the first row we obtain

$$|\,\lambda I \;-\; A\,| \;=\; f(\lambda) \;=\; (\lambda \;-\; a_{11})C_{11} \;+\; \Sigma_{j=2}^n\,(-a_{1j}\,)C_{1j} \tag{68}$$

Here, C_{1j} is the cofactor of a_{1j} in the matrix $|\lambda I - A|$. Note that there are only n-2 elements involving a_{ii} - λ in the matrices C_{12}, C_{13} ..., C_{1n}. Given that there are only n-2 elements involving

a_{ii} - λ in the matrices C_{1j}, ...,C_{1n}, when we later expand these matrices, there will be no powers of λ greater than n-2. This then means that we can write 68 as

$$|\,\lambda I \,-\, A\,| \,=\, (\lambda \,-\, a_{11})C_{11} \,+ [terms\; of\; degree\; n-2\; or\; less\; in\; \lambda] \tag{69}$$

We can then make similar arguments about C_{11} and C_{22} and so on and thus obtain

$$|\,\lambda I \,-\, A\,| \,=\, (\lambda \,-\, a_{11})(\lambda \,-\, a_{22}) \ldots (\lambda \,-\, a_{nn}) \,+\, [\text{ terms of degree n-2 or less in } \lambda\,] \tag{70}$$

Consider the first term in equation 70. It is the same as equation 34 with a_{11} replacing λ_i. We can therefore write this first term as follows

$$\begin{aligned}(\lambda \,-\, a_{11})(\lambda \,-\, a_{22}) \ldots (\lambda \,-\, a_{nn}) \,&=\lambda^n -\, \lambda^{n-1}\, \Sigma_{i=1}^n\, a_{ii} \\ &+ \lambda^{n-2}\, \Sigma_{i\neq j}\; a_{ii}\, a_{jj} \\ &- \lambda^{n-3}\, \Sigma_{i\neq j\neq k} a_{ii} a_{jj} a_{kk} \\ &+ \lambda^{n-4}\, \Sigma_{i\neq j\neq k\neq \ell} a_{ii} a_{jj} a_{kk} a_{\ell\ell} \\ &+ \cdots \\ &+ (-1)^n\, \Pi_{i=1}^n\, a_{ii}\end{aligned} \tag{71}$$

Substituting in equation 70, we obtain

$$\begin{aligned}|\,\lambda I \,-\, A\,| \,&= (\lambda \,-\, a_{11})(\lambda \,-\, a_{22}) \ldots (\lambda \,-\, a_{nn}) \,+\, [terms\; of\; degree\; n-2\; or\; less\; in\; \lambda] \\ &= \lambda^n -\, \lambda^{n-1}\, \Sigma_{i=1}^n\, a_{ii} \,+\, \lambda^{n-2}\, \Sigma_{i\neq j}\; a_{ii}\, a_{jj} \\ &- \lambda^{n-3}\, \Sigma_{i\neq j\neq k} a_{ii} a_{jj} a_{kk} \,+\, \lambda^{n-4}\, \Sigma_{i\neq j\neq k\neq \ell} a_{ii} a_{jj} a_{kk} a_{\ell\ell} \,+\, \cdots \\ &+ (-1)^n\, \Pi_{i=1}^n\, a_{ii} \,+\, [\text{other terms of degree n-2 or less in } \lambda\,]\end{aligned} \tag{72}$$

Now the compare the coefficient of $(-\,\lambda^{n-1})$ in equation 72 with that in equation 36b or 60b which we repeat here for convenience.

$$|\,\lambda I - A\,| \,= \lambda^n \quad -\lambda^{n-1}\, \Sigma_{i=1}^{n-1}\, \lambda_i \quad +\lambda^{n-2}\, \Sigma_{i\neq j} \lambda_i \lambda_j \quad - \cdots \quad +(-1)^n\, \Pi_{i=1}^n\, \lambda_i \tag{73}$$

It is clear from this comparison that

$$\Sigma_{i=1}^n\, a_{ii} \,=\, \Sigma_{i=1}^n\, \lambda_i \tag{74}$$

□

Note that this expansion provides an alternative proof of theorem 7.

6. Transposes and inverses

6.1. **Statement of theorem.**

Theorem 9. *Let A be a square matrix of order n with characteristic roots λ_i, i = 1,...,n. Then*

- **1:** *The characteristic roots of A' are the same as those of A.*
- **2:** *If A is nonsingular, the characteristic roots of A^{-1} are given by $\mu_i = \frac{1}{\lambda_i}$, i = 1,...,n..*

6.2. **proof of part 1.**

Proof. The characteristic roots of A and A' are the solutions of the equations

$$|\,\lambda I \,-\, A\,| \,=0 \tag{75a}$$

$$|\,\theta I \,-\, A'\,| \,=0 \tag{75b}$$

Now note that $\theta I - A' = (\theta I - A)'$. Furthermore it is clear from the definition of a determinant that $|B| = |B'|$ since the permutations could as well be taken over rows as columns. This means then that $|(\theta I - A)'| = |\theta I - A|$ which implies that $\theta_i = \lambda_i$

□

6.3. **proof of part 2.**

Proof. Write the characteristic equation of A^{-1} as

$$|\,\mu I \,-\, A^{-1}\,| \,=\, 0 \tag{76}$$

Now rewrite the matrix μI - A^{-1} in the following fashion

$$\begin{aligned}
\mu I \,-\, A^{-1} &= A^{-1}(\mu A \,-\, I) && \text{postmultiply each term by A and then premultiply} \\
& && \text{the whole expression by } A^{-1} \\
&= (-\mu)\; A^{-1}(-\frac{1}{\mu}\,\mu A \,+\, \frac{1}{\mu}\; I) && \text{multiply each term inside the parentheses by } \left(\frac{-1}{\mu}\right) \\
& && \text{and premultiply whole expression by } (-\mu) \\
&= (-\mu)\; A^{-1}(-A \,+\, \frac{1}{\mu}\; I) && \text{simplify} \\
&= (-\mu)\; A^{-1}(\frac{1}{\mu}\; I \,-\, A) && \text{commute}
\end{aligned} \tag{77}$$

Now take the determinant of both sides of equation 77

$$\begin{aligned}
|\,\mu I \,-\, A^{-1}\,| \;&=\; \left|-\mu A^{-1}\,\left(\frac{1}{\mu}I \,-\, A\right)\right| \\
&= \left|-\mu A^{-1}\,|\,|\,\left(\frac{1}{\mu}I \,-\, A\right)\right| \\
&= (-1)^n\,\mu^n\;|\,A^{-1}\,|\,|\,\lambda I \,-\, A\,|, \quad \text{where } \lambda \,=\, \frac{1}{\mu}
\end{aligned} \tag{78}$$

If μ = 0, then $\left|\,\mu I \,-\, A^{-1}\,\right|$ = $\left|\,-\,A^{-1}\,\right|$ = $(-1)^n\;\left|\,A^{-1}\,\right|$. Because the matrix is A nonsingular, A^{-1} is also non-singular. This imples that μ cannot equal zero, so μ = 0 is not a root of the equation. Thus the expression on the right hand side of 78 can be zero iff $|\lambda I - A|$ = 0, where $\lambda = \frac{1}{\mu}$. Thus if μ_i are the roots of A^{-1} we must have that

$$\mu_j \;=\; \frac{1}{\lambda_j}, \quad j \,=\, 1,\, 2,\, \ldots\, ,\, n \tag{79}$$

□

7. Orthogonal matrices

7.1. Orthogonal and orthonormal vectors and matrices.

Definition 1. The nx1 vectors a and b are orthogonal if

$$a'b = 0 \tag{80}$$

Definition 2. The nx1 vectors a and b are orthonormal if

$$\begin{aligned} a'b &= 0 \\ a'a &= 1 \\ b'b &= 1 \end{aligned} \tag{81}$$

Definition 3. The square matrix Q (nxn) is orthogonal if its columns are orthonormal. Specifically

$$\begin{aligned} Q &= \begin{pmatrix} x^1 & x^2 & x^3 & \cdots & x^n \end{pmatrix} \\ x'^i\, x^i &= 1 \\ x'^i\, x^j &= 0 \; i \neq j \end{aligned} \tag{82}$$

A consequence of the definition is that

$$Q'Q = I \tag{83}$$

7.2. Non-singularity property of orthogonal matrices.

Theorem 10. *An orthogonal matrix is nonsingular.*

Proof. If the columns of the matrix are independent then it is nonsingular. Consider the matrix Q given by

$$Q = [q^1 \; q^2 \; \cdots \; q^n] = \begin{bmatrix} q_1^1 & q_1^2 & \cdots & q_1^n \\ q_2^1 & q_2^2 & \cdots & q_2^n \\ q_3^1 & q_3^2 & \cdots & q_3^n \\ \vdots & \vdots & \vdots & \vdots \\ q_n^1 & q_n^2 & \cdots & q_n^n \end{bmatrix} \tag{84}$$

If the columns are dependent then,

$$a_1 \begin{pmatrix} q_1^1 \\ q_2^1 \\ \vdots \\ q_n^1 \end{pmatrix} + a_2 \begin{pmatrix} q_1^2 \\ q_2^2 \\ \vdots \\ q_n^2 \end{pmatrix} + a_3 \begin{pmatrix} q_1^3 \\ q_2^3 \\ \vdots \\ q_n^3 \end{pmatrix} + \cdots + a_n \begin{pmatrix} q_1^n \\ q_2^n \\ \vdots \\ q_n^n \end{pmatrix} = 0, \quad a_i \neq 0 \text{ for at least one i} \tag{85}$$

Now premultiply the dependence equation by the transpose of one of the columns of Q, say q^j.

$$
\begin{aligned}
& a_1 \, [q_1^j \quad q_2^j \quad \cdots \quad q_n^j] \begin{pmatrix} q_1^1 \\ q_2^1 \\ \vdots \\ q_n^1 \end{pmatrix} \, + \, a_2 \, [q_1^j \quad q_2^j \quad \cdots \quad q_n^j] \begin{pmatrix} q_1^2 \\ q_2^2 \\ \vdots \\ q_n^2 \end{pmatrix} \\
& + \, \cdots \, + \, a_n \, [q_1^j \quad q_2^j \quad \cdots \quad q_n^j] \begin{pmatrix} q_1^n \\ q_2^n \\ \vdots \\ q_n^n \end{pmatrix} \, = \, 0 \, , \, a_i \, \neq \, 0 \; for \; at \; least \; one \; i
\end{aligned}
\tag{86}
$$

By orthogonality all the terms involving j and i ≠ j are zero. This then implies

$$
a_j \, [q_1^j \quad q_2^j \quad \cdots \quad q_n^j] \begin{pmatrix} q_1^j \\ q_2^j \\ \vdots \\ q_n^j \end{pmatrix} \, = \, 0 \tag{87}
$$

But because the columns are orthonormal $q'^j \, q^j = 1$ which implies that $a^j = 0$. Because j is arbitrary this implies that all the a^j are zero. Thus the columns are independent and the matrix has an inverse.

□

7.3. **Tranposes and inverses of orthogonal matrices.**

Theorem 11. *If Q is an orthogonal matrix, then $Q' = Q^{-1}$.*

Proof. By the definition of an orthogonal matrix $Q'Q$ = I. Now postmultiply the identity by Q^{-1}. It exists by Theorem 10.

$$
\begin{aligned}
Q' \, Q \; &= I \\
Q' \, Q \, Q^{-1} \; &= I Q^{-1} \\
\Rightarrow \; Q' \; &= Q^{-1}
\end{aligned}
\tag{88}
$$

□

7.4. **Determinants and characteristic roots of orthogonal matrices.**

Theorem 12. *If Q is an orthogonal matrix of order n ,then*

a: $|Q| = 1$ *or* $|Q| = -1$

b: *If λ_i is a characteristic root of Q, then $\lambda_i = \pm 1$, i=1,...,n.*

Proof. To show note that $Q'Q$ = I. Then

$$
\begin{aligned}
1 \; &= | \, I \, | \\
&= | \, Q' \, Q \, | \; = \; | \, Q' \, | \, | \, Q \, | \\
&= | \, Q \, |^2 \quad \text{because} \quad |Q| \; = \; |Q'\,|
\end{aligned}
\tag{89}
$$

To show the second part note that $Q' = Q^{-1}$. Also note that the characteristic roots of Q are the same as the roots of Q'. By theorem 9, the characteristic roots of Q^{-1} are $(1/\lambda_i)$. This implies then that the roots of Q' are the same as Q^{-1}. This means then that $\lambda_i = 1/\lambda_i$. This can be true iff $\lambda_i = \pm 1$.

□

8. A Digression on Complex Numbers

8.1. **Definition of a complex number.** A complex number is an ordered pair of real numbers denoted by (x_1, x_2). The first member, x_1, is called the real part of the complex number; the second member, x_2, is called the imaginary part. We define equality, addition, subtraction, and multiplication so as to preserve the familiar rules of algebra for real numbers.

8.1.1. *Equality of complex numbers.* Two complex numbers x = (x_1, x_2) and y = (y$_1$, y_2) are called equal iff

$$x_1 = y_1, \text{ and } x_2 = y_2. \tag{90}$$

8.1.2. *Sum of complex numbers.* The sum of two complex numbers x + y is defined as

$$x + y = (x_1 + y_1, x_2 + y_2) \tag{91}$$

8.1.3. *Difference of complex numbers.* To subtract two complex numbers, the following rule applies.

$$x - y = (x_1 - y_1, x_2 - y_2) \tag{92}$$

8.1.4. *Product of complex numbers.* The product xy is defined as

$$xy = (x_1\, y_1 - x_2\, y_2, x_1\, y_2 + x_2\, y_1). \tag{93}$$

The properties of addition and multiplication defined satisfy the commutative, associative and distributive laws.

8.2. **The imaginary unit.** The complex number (0,1) is denoted by i and is called the imaginary unit. We can show that i^2 = -1 as follows.

$$i^2 = (0, 1)(0, 1) = (0 - 1, 0 + 0) = (-1, 0) = -1 \tag{94}$$

8.3. **Representation of a complex number.** A complex number x = (x_1, x_2) can be written in the form

$$x = x_1 + i\, x_2. \tag{95}$$

Alternatively a complex number z = (x,y) is sometimes written

$$z = x + iy. \tag{96}$$

8.4. **Modulus of a complex number.** The modulus or absolute value of a complex number x = (x_1, x_2) is the nonnegative real number $|x|$ given by

$$|x| = \sqrt{x_1^2 \, + \, x_2^2} \tag{97}$$

8.5. **Complex conjugate of a complex number.** For each complex number z = x + iy, the number z_{-} = x - iy is called the complex conjugate of z. The product of a complex number and its conjugate is a real number. In particular, if z = x + iy then

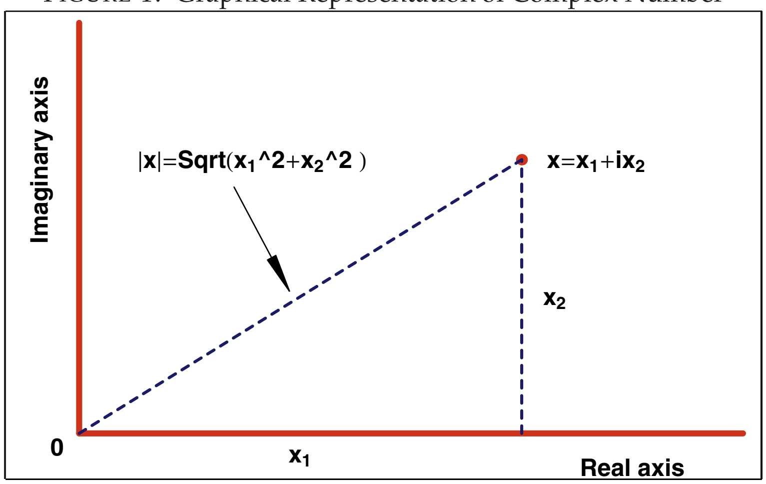
$$z\,z_{-} = (x, y)(x, -y) = (x^2 + y^2, -xy + yx) = (x^2 + y^2, 0) = x^2 + y^2 \tag{98}$$

Sometimes we will use the notation $\bar{z}$ to represent the complex conjugate of a complex number. So $\bar{z}$ = x - iy. We can then write

$$z\,\bar{z} = (x, y)(x, -y) = (x^2 + y^2, -xy + yx) = (x^2 + y^2, 0) = x^2 + y^2 \tag{99}$$

8.6. **Graphical representation of a complex number.** Consider representing a complex number in a two dimensional graph with the vertical axis representing the imaginary part. In this framework the modulus of the complex number is the distance from the origin to the point. This is seen clearly in figure 1

FIGURE 1. Graphical Representation of Complex Number



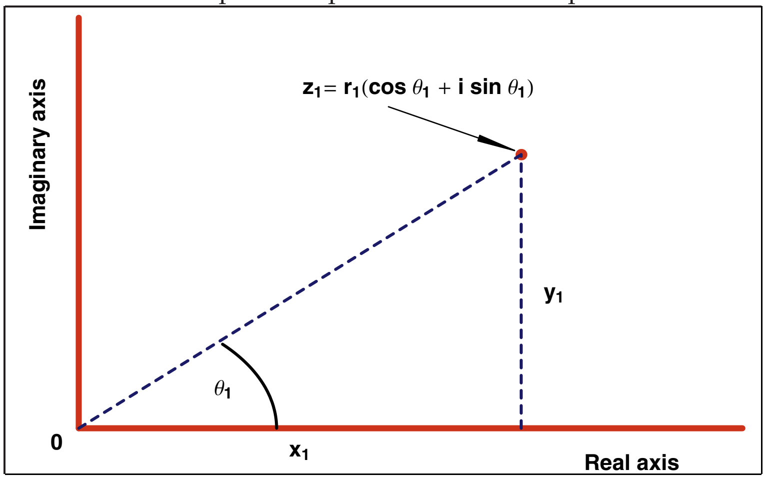
8.7. **Polar form of a complex number.** We can represent a complex number by its angle and distance from the origin. Consider a complex number $z_1 = x_1 + \mathrm{i}\, y_1$. Now consider the angle θ_1 which the ray from the origin to the point z_1 makes with the x axis. Let the modulus of z be denoted by r_1. Then cos $\theta_1 = x_1/r_1$ and sin $\theta_1 = y_1/r_1$. This then implies that

$$\begin{aligned} z_1 &= x_1 + i\,y_1 = r_1 \cos\theta_1 + i\,r_1 \sin\theta_1 \\ &= r_1 (\cos\theta_1 + i \sin\theta_1) \end{aligned} \tag{100}$$

Figure 2 shows how a complex number is represented in polar corrdinates.

8.8. **Complex Exponentials.** The exponential e^x is a real number. We want to define e^z when z is a complex number in such a way that the principle properties of the real exponential function will be preserved. The main properties of e^x, for x real, are the law of exponents, $e^{x_1}e^{x_2} = e^{x_1+x_2}$ and the equation $e^0 = 1$. If we want the law of exponents to hold for complex numbers, then it must be that

FIGURE 2. Graphical Representation of Complex Number



$$e^z \;=\; e^{x+iy} \;=\; e^x\, e^{iy} \tag{101}$$

We already know the meaning of e^x. We therefore need to define what we mean by e^{iy}. Specifically we define e^{iy} in equation 102

Definition 4.

$$e^{iy} \;=\; \cos\, y + i\, \sin\, y \tag{102}$$

With this in mind we can then define $e^z = e^{x+iy}$ as follows

Definition 5.

$$e^z \;=\; e^x\, e^{iy} \;=\; e^x(\cos\, y + i\, \sin\, y) \tag{103}$$

Obviously if x = 0 so that z is a pure imaginary number this yields

$$e^{iy} \;=\; (\cos\, y + i\, \sin\, y) \tag{104}$$

It is easy to show that $e^0 = 1$. If z is real then y = 0. Equation 105 then becomes

$$\begin{aligned} e^z \; &= e^x\, e^{i\,0} \;=\; e^x(\cos\,(0) + i\, \sin\,(0)\,) \\ &= e^x\, e^{i\,0} \;=\; e^x(1 \;+\; 0\,) \;=\; e^x \end{aligned} \tag{105}$$

So e^0 obviously is equal to 1.

To show that $e^x e^{iy} = e^{x\,+\,iy}$ or $e^{z_1} e^{z_2} = e^{z_1\,+\,z_2}$ we will need to remember some trigonometric formulas.

Theorem 13.

$$\sin(\phi \;+\; \theta) \;=\; \sin\,\phi\,\cos\,\theta \;+\; \cos\,\phi\,\sin\,\theta \tag{106a}$$

$$\sin(\phi \;-\; \theta) \;=\; \sin\,\phi\,\cos\,\theta \;-\; \cos\,\phi\,\sin\,\theta \tag{106b}$$

$$\cos(\phi \;+\; \theta) \;=\; \cos\,\phi\,\cos\,\theta \;-\; \sin\,\phi\,\sin\,\theta \tag{106c}$$

$$\cos(\phi \;-\; \theta) \;=\; \cos\,\phi\,\cos\,\theta \;+\; \sin\,\phi\,\sin\,\theta \tag{106d}$$

Now to the theorem showing that $e^{z_1} e^{z_2} = e^{z_1 + z_2}$

Theorem 14. *If $z_1 = x_1 + iy_t$ and $z_2 = x_2 + iy_2$ are two complex numbers, then we have $e^{z_1} e^{z_2} = e^{z_1+z_2}$.*

Proof.

$$\begin{aligned} e^{z_1} &= e^{x_1}(cos\ y_1 + i\ sin\ y_1),\ e^{z_2} = e^{x_2}(cos\ y_1 + i\ sin\ y_2), \\ e^{z_1} e^{z_2} &= e^{x_1} e^{x_2}[cos\ y_1\ cos\ y_2 - sin\ y_1\ sin\ y_2 \\ &\quad + i(cos\ y_1\ sin\ y_2 + sin\ y_1\ cos\ y_2)]. \end{aligned} \tag{107}$$

Now $e^{x_1} e^{x_2} = e^{x_1+x_2}$, since x_1 and x_2 are both real. Also,

$$cos\ y_1\ cos\ y_2 - sin\ y_1\ sin\ y_2 = cos\,(y_1 + y_2) \tag{108}$$

and

$$cos\ y_1\ sin\ y_2 + sin\ y_1\ cos\ y_2 = sin\,(y_1 + y_2), \tag{109}$$

and hence

$$e^{z_1} e^{z_2} = e^{x_1+x_2}[cos\,(y_1 + y_2) + i\ sin\,(y_1 + y_2) = e^{z_1+z_2}. \tag{110}$$

□

Now combine the results in equations 100, 105 and 104 to obtain the polar representation of a complex number

$$\begin{aligned} z_1 &= x_1 + i\,y_1 \\ &= r_1\,(\cos\ \theta_1 + i\ \sin\ \theta_1) \\ &= r_1\,e^{i\,\theta_1}, \qquad \text{(by equation 102)} \end{aligned} \tag{111}$$

The usual rules for multiplication and division hold so that

$$\begin{aligned} z_1\ z_2 &= r_1\,e^{i\theta_1} r_2\,e^{i\theta_2} = (r_1\,r_2)\,e^{i(\theta_1+\ \theta_2)} \\ \frac{z_1}{z_2} &= \frac{r_1\,e^{i\theta_1}}{r_2\,e^{i\theta_2}} = \left(\frac{r_1}{r_2}\right)\ e^{i(\theta_1-\ \theta_2)} \end{aligned} \tag{112}$$

8.9. **Complex vectors.** We can extend the concept of a complex number to complex vectors where z = x + iy is a vector as are the components x and y. In this case the modulus is given by

$$|z| = \sqrt{x'x + y'y} = (x'x + y'y)^{\frac{1}{2}} \tag{113}$$

9. Symmetric matrices and characteristic roots

9.1. Symmetric matrices and real characteristic roots.

Theorem 15. *If S is a symmetric matrix whose elements are real, its characteristic roots are real.*

Proof. Let $\lambda = \lambda_1 + i\lambda_2$ be a characteristic root of S and let z = x + iy be its associated characteristic vector. By definition of a characteristic root

$$\begin{aligned} Sz &= \lambda z \\ S(x + iy) &= (\lambda_1 + i\lambda_2)(x + iy) \end{aligned} \tag{114}$$

Now premultiply 114 by the complex conjugate (written as a row vector) of z denoted by $\bar{z}'$. This will give

$$\begin{aligned} \bar{z}' S z &= \lambda \bar{z}' z \\ (x - iy)' S (x + iy) &= (\lambda_1 + i\lambda_2)(x - iy)'(x + iy) \end{aligned} \tag{115}$$

We can simplify 115 as follows

$$\begin{aligned} (x - iy)' S (x + iy) &= (\lambda_1 + i\lambda_2)(x - iy)'(x + iy) \\ \Rightarrow (x'S - y'iS)(x + iy) &= (\lambda_1 + i\lambda_2)\,(x'x + y'y, x'y - y'x) \\ \Rightarrow x'Sx + ix'Sy - iy'Sx - (i\,i)y'Sy &= (\lambda_1 + i\lambda_2)\,(x'x + y'y, 0) \\ \Rightarrow x'Sx + iy'S'x - iy'Sx - (i\,i)y'Sy &= (\lambda_1 + i\lambda_2)\,(x'x + y'y, 0), \quad x'Sy \text{ a scalar} \\ \Rightarrow x'Sx + iy'Sy - iy'Sx - (i\,i)y'Sy &= (\lambda_1 + i\lambda_2)\,(x'x + y'y, 0), \quad \text{S is symmetric} \\ \Rightarrow x'Sx + y'Sy &= (\lambda_1 + i\lambda_2)\,(x'x + y'y, 0), \quad \text{i i = -1} \\ \Rightarrow x'Sx + y'Sy &= (\lambda_1 + i\lambda_2)\,(x'x + y'y) \end{aligned} \tag{116}$$

Further note that

$$\begin{aligned} \bar{z}'z &= (x - iy)'(x + iy) \\ &= (x'x + y'y, x'y - y'x) \\ &= (x'x + y'y, 0) \\ &= x'x \; + \; y'y \; > \; 0 \end{aligned} \tag{117}$$

is a real number. The left hand side of 116 is a real number. Given that $x'x \; + \; y'y$ is positive and real, this implies that $\lambda_2 = 0$. This then implies that $\lambda = \lambda_1 + i\lambda_2 = \lambda_1$ and is real.

□

Theorem 16. *If S is a symmetric matrix whose elements are real, corresponding to any characteristic root there exist characteristic vectors that are real.*

Proof. The matrix S is symmetric so we can write equation 114 as follows

$$\begin{aligned} Sz &= \lambda z \\ S(x + iy) &= (\lambda_1 + i\lambda_2)(x + iy) \\ &= (\lambda_1)(x + iy) \\ &= \lambda_1 x + i\lambda_1 y \end{aligned} \tag{118}$$

Thus $z = x + iy$ will be a characteristic vector of S corrresponding to $\lambda = \lambda_1$ as long as x and y satisfy

$$\begin{aligned} Ax &= \lambda x \\ Ay &= \lambda y \end{aligned} \tag{119}$$

and both are not zero, so that $z \neq 0$. Then construct a real characteristic vector by choosing $x \neq 0$, such that $Ax = \lambda_1 x$ and y = 0.

□

9.2. **Symmetric Matices and Orthogonal Characteristic Vectors.**

Theorem 17. *If A is a symmetric matrix, the characteristic vectors corresponding to different characteristic roots are orthogonal, i.e., if x^i corresponds to λ_i and x^j corresponds to λ_j ($\lambda_i \neq \lambda_j$), then ${x^i}'x^j = 0$.*

Proof. We will use the fact that the matrix is symmetric and characteristic root equation 1. Multiply the equations by ${x^j}'$ and ${x^i}'$

$$\begin{aligned} Ax^i &= \lambda_i x^i \\ \Rightarrow\ {x^j}'Ax^i &= \lambda_i {x^j}'x^i \qquad (120a) \\ Ax^j &= \lambda_j x^j \\ \Rightarrow\ {x^i}'Ax^j &= \lambda_j {x^i}'x^j \qquad (120b) \end{aligned}$$

Because the matrix A is symmetric, it has real characteristic vectors. The inner product of two of them will be a real scalar, so

$${x^j}'x^i = {x^i}'x^j \qquad (121)$$

Now subtract 120b for 120a using 121

$$\begin{aligned} {x^j}'Ax^i &= \lambda_i {x^j}'x^i \\ -\ {x^i}'Ax^j &= \lambda_j {x^i}'x^j \\ ---\quad & \quad --- \\ {x^j}'Ax^i - {x^i}'Ax^j &= (\lambda_i - \lambda_j){x^j}'x^i \end{aligned} \qquad (122)$$

Because A is symmetric and the characteristic vectors are real, ${x^j}'Ax^i$ is symmetric and real. Therefore ${x^j}'Ax^i = {x^i}'A'x^j = {x^i}'Ax^j$. The left hand side of 122 is therfore zero. This then implies

$$\begin{aligned} 0 &= (\lambda_i - \lambda_j){x^j}'x^i \\ \Rightarrow\ {x^j}'x^i &= 0, \qquad \lambda_i \neq \lambda_j \end{aligned} \qquad (123)$$

□

9.3. **An Aside on the Row Space, Column Space and the Null Space of a Square Matrix.**

9.3.1. *Definition of R^n.* The space R^n consists of all column vectors with n components. The components are real numbers.

9.3.2. *Subspaces.* A subspace of a vector space is a set of vectors (including 0) that satisfies two requirements: If a and b are vectors in the subspace and c is any scalar, then

1: a + b is in the subspace
2: ca is in the subspace

9.3.3. *Subspaces and linear combinations.* A subspace containing the vectors a and b must contain all linear combinations of a and b.

9.3.4. *Column space of a matrix.* The column space of the mxn matrix A consists of all linear combinations of the columns of A. The combinations can be written as Ax. The column space of A is a subspace of R^m because each of the columns of A has m rows. The system of equations Ax= b is solvable if and only if b is in the column space of A. What this means is that the system is solvable if there is some way to write b as a linear combination of the columns of A.

9.3.5. *Nullspace of a matrix.* The nullspace of an m$\times$n matrix A consists of all solutions to Ax = 0. These vectors x are in R^n. The elements of x are the multipliers of the columns of A whose weighted sum gives the zero vector. The nullspace containing the solutions **x** is denoted by **N**(A). The null space is a subspace of R^n while the column space is a subspace of R^m. For many matrices, the only solution to Ax = 0 is x = 0. If n > m, the nullspace will always contain vectors other than x = 0.

9.3.6. *Basis vectors.* A set of vectors in a vector space is a **basis** for that vector space if any vector in the vector space can be written as a linear combination of them. The minimum number of vectors needed to form a basis for R^k is k. For example, in R^2, we need two vectors of length two to form a basis. One vector would only allow for other points in R^2 that lie along the line through that vector.

9.3.7. *Linear dependence.* A set of vectors is linearly dependent if any one of the vectors in the set can be written as a linear combination of the others. The largest number of linearly independent vectors we can have in R^k is k.

9.3.8. *Linear independence.* A set of vectors is linearly independent if and only if the only solution to

$$\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_{k\, a_k} = 0 \tag{124}$$

is

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 \tag{125}$$

We can also write this in matrix form. The columns of the matrix A are independent if the only solution to the equation Ax = 0 is x = 0.

9.3.9. *Spanning vectors.* The set of all linear combinations of a set of vectors is the vector space spanned by those vectors. By this we mean all vectors in this space can be written as a linear combination of this particular set of vectors.

9.3.10. *Rowspace of a matrix.* The rowspace of the $m \times n$ matrix A consists of all linear combinations of the rows of A. The combinations can be written as x′A or A′x depending on whether one considers the resulting vectors to be rows or columns. The row space of A is a subspace of R^n because each of the rows of A has n columns. The **row space** of a matrix is the subspace of R^n spanned by the rows.

9.3.11. *Linear independence and the basis for a vector space.* A basis for a vector space with k dimensions is any set of k linearly independent vectors in that space

9.3.12. *Formal relationship between a vector space and its basis vectors.* A **basis** for a vector space is a sequence of vectors that has two properties simultaneously.

1: The vectors are linearly independent
2: The vectors span the space

There will be one and only one way to write any vector in a given vector space as a linear combination of a set of basis vectors. There are an infinite number of basis vectors for a given space, but only one way to write any given vector as a linear combination of a particular basis.

9.3.13. *Bases and Invertible Matrices.* The vectors $\xi_1, \xi_2, \ldots, \xi_n$ are a basis for R^n exactly when they are the columns of and $n \times n$ invertible natirx. Therefore R^n has infinitely many bases, one associated with every differnt invertible matrix.

9.3.14. *Pivots and Bases.* When reducing an $m \times n$ matrix A to row-echelon form, the pivot columns form a basis for the column space of the matrix A. The the pivot rows form a basis for the row space of the matrix A.

9.3.15. *Dimension of a vector space.* The dimension of a vector space is the number of vectors in every basis. For example, the dimension of R^2 is 2, while the dimension of the vector space consisting of points on a particular plane is R^3 is also 2.

9.3.16. *Dimension of a subspace of a vector space.* The dimension of a subspace space S_n of an n-dimensional vector space V_n is the maximum number of linearly independent vectors in the subspace.

9.3.17. *Rank of a Matrix.*

1: The number of non-zero rows in the row echelon form of an $m \times n$ matrix A produced by elementary operations on A is called the rank of A.
2: The rank of an $m \times n$ matrix A is the number of pivot columns in the row echelon form of A.
3: The column rank of an $m \times n$ matrix A is the maximum number of linearly independent columns in A.
4: The row rank of an $m \times n$ matrix A is the maximum number of linearly independent rows in A
5: The column rank of an $m \times n$ matrix A is eqaul to row rank of the $m \times n$ matrix A. This common number is called the **rank** of A.
6: An $n \times n$ matrix A with rank = n is said to be of full rank.

9.3.18. *Dimension and Rank.* The dimension of the column space of an $m \times n$ matrix A equals the rank of A, which also equals the dimension of the row space of A. The *number of independent columns* of A *equals the number of independent rows* of A. As stated earlier, the r columns containing pivots in the row echelon form of the matrix A form a basis for the cloumn space of A.

9.3.19. *Vector spaces and matrices.* An mxn matrix A with full column rank (i.e., the rank of the matrix is equal to the number of columns) has all the following properties.

1: The n columns are independent
2: The only solution to AX = 0 is x = 0.
3: The rank of the matrix = dimension of the column space = n
4: The columns are a basis for the column space.

9.3.20. *Determinants, Minors, and Rank.*

Theorem 18. *The rank of an $m \times n$ matrix A is k if and only if every minor in A of order $k+1$ vanishes, while there is at least one minor of order k which does not vanish.*

Proposition 1. Consider an $m \times n$ matrix A.

1: det A= 0 if every minor of order n - 1 vanishes.
2: If every minor of order n equals zero, then the same holds for the minors of higher order.
3 (restatement of theorem): The largest among the orders of the non-zero minors generated by a matrix is the rank of the matrix.

9.3.21. *Nullity of an $m \times n$ Matrix A.* The nullspace of an $m \times n$ Matrix A is made up of vectors in R^n and is a subspace of R^n. The dimension of the nullspace of A is called the nullity of A. It is the the maximum number of linearly independent vectors in the nullspace.

9.3.22. *Dimension of Row Space and Null Space of an $m \times n$ Matrix A.* Consider an $m \times n$ Matrix A. The dimension of the nullspace, the maximum number of linearly independent vectors in the nullspace, plus the rank of A, is equal to n. Specifically,

Theorem 19.

$$rank(A) + nullity(A) = n \tag{126}$$

9.4. **Gram-Schmidt Orthogonalization and Orthonormal Vectors.**

Theorem 20. *Let e^i, i= 1,2,...,n be a set of n linearly independent n-element vectors. They can be transformed into a set of orthonormal vectors.*

Proof. Transform them into an orthogonal set and divide each vector by the square root of its length. Start by defining the orthogonal vectors as

$$\begin{aligned} y^1 &= e^1 \\ y^2 &= a_{12}e^1 \ + \ e^2 \\ y^3 &= a_{13}e^1 \ + \ a_{23}e^2 \ + \ e^3 \\ &\ \vdots \\ y^n &= a_{1n}e^1 \ + \ a_{2n}e^2 \ + \ \cdots + a_{n-1}e^{n-1} \ + \ e^n \end{aligned} \tag{127}$$

The a_{ij} must be chosen in a way such that the y vectors are orthogonal

$$y^{i\prime} y^j \ = \ 0, \quad i = 1, 2, ..., j - 1 \tag{128}$$

Note that since $y^{i\prime}$ depends only on the e^i, the following condition is equivalent

$$e^{i\prime} y^j \ = \ 0, \quad i = 1, 2, ..., j - 1 \tag{129}$$

Now define matrices containing the above information.

$$\begin{aligned} X_j &= [e^1 \ e^2 \ e^3 \ \cdots \ e^{j-1}] \\ a^j &= \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{(j-1)\,j} \end{pmatrix} \end{aligned} \tag{130}$$

The matrix $X_2 = [e_1]$, $X_3 = [e_1 \ e_2]$, $X_4 = [e_1\, e_2\, e_3]$ and so forth. Now rewrite the system of equations defining the a_{ij}.

$$\begin{aligned} y^1 &= e^1 \\ y^j &= X_j a^j \ + \ e^j \\ &= [e^1 \ e^2 \ e^3 \ \cdots \ e^{j-1}] \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{(j-1)\,j} \end{pmatrix} \ + \ e^j \\ &= a_{1j}e^1 \ + \ a_{2j}e^2 + ... + a_{(j-1)\,j}e^{j-1} \ + \ e^j \end{aligned} \tag{131}$$

We can write the condition that $y^{i\prime}$ and $y^{j\prime}$ be orthogonal for all i and j as follows

$$\begin{aligned} e^{i\prime} y^j &= 0, \quad i = 1, 2, ..., j - 1 \\ \Rightarrow \ X_j{}' y^j &= 0 \\ \Rightarrow \ X_j{}' X_j a^j \ + \ X_j{}' X_j \, e^j &= 0, \qquad \text{(substitute from 131)} \end{aligned} \tag{132}$$

The columns of X_j are independent. This implies that the matrix ($X_j{}'X_j$) can be inverted. Then we can solve the system as follows

$$\begin{aligned} X_j{}'X_j a^j \; + \; X_j{}'X_j \, e^j \; &= 0 \\ \Rightarrow X_j{}'X_j a^j \; &= \; - \, X_j{}'e^j \\ \Rightarrow \; a^j \; &= \; -(X_j{}'X_j)^{-1} X_j{}'e^j \end{aligned} \tag{133}$$

The system in equation 131 can now be written

$$\begin{aligned} y^1 \; &= e^1 \\ y^j \; &= - \; X_j (X_j{}'X_j)^{-1} X_j{}'e^j \; + \; e^j \\ y^j \; &= e^j \; - \; X_j (X_j{}'X_j)^{-1} X_j{}'e^j \end{aligned} \tag{134}$$

Normalizing gives orthonormal vectors v

$$v^i \; = \; \frac{y^i}{(y^i \, y^i)^{1/2}} \; i \; = \; 1, 2, ..., n \tag{135}$$

The vectors $y^{i\prime}$ and $y^{j\prime}$ are orthogonal by definition. We can see that they are orthonormal as follows

$$\begin{aligned} v^{i\prime} v^i \; &= \frac{y^{i\prime}}{(y^{i\prime} \, y^i)^{1/2}} \frac{y^i}{(y^{i\prime} \, y^i)^{1/2}} \\ &= \frac{y^{i\prime} y^i}{(y^{i\prime} \, y^i)} \; = \; 1 \end{aligned} \tag{136}$$

□

9.5. **Symmetric Matices and Linearly Independent Orthonormal Characteristic Vectors.**

Theorem 21. *Let A be an nxn symmetric matrix with possibly non-distinct characteristic roots. Let the distinct roots be* $\lambda_i, \; i = 1, 2, \ldots, s \leq n$ *and let the multiplicty of j*$^{\text{th}}$ *root be* k_j *with* $\sum_{k=1}^{s} k_j \; = \; n$*. Then*

- **1:** *If a characteristic root* λ_j *has multiplicity* $k_j \geq 2$*, there exist* k_j *orthonormal linearly independent characteristic vectors with characteristic root* λ_j*.*
- **2:** *There cannot be more than* k_j *linearly independent characteristic vectors with the same characteristic root* λ_j*, hence, if a characteristic root has multiplicity* k_j*, the characteristic vectors with root* k_j *span a subspace of* R^n *with dimension* k_j*. Thus the orthonormal set in [1] spans a* k_j *dimensional subspace of* R^n*. By combining these bases for different subspaces, a basis for the original space* R^n *can be obtained.*

Proof. Let λ_j be a characteristic root of A. Let its associated characteristic vector be given by q^1. Assume that the multiplicity of q^1 is k_1. Assume that it is normalized to length 1. Now choose a set of n-1 vectors u (of length n) that are linearly independent of each other and q^1. Number them u^2, ..., u^k. Use the Gram-Schmidt theorem to form an orthonormal set of vectors. Notice that the algorithm will not change q^1 because it is the first vector. Call this matrix of vectors Q_1. Specifically

$$Q_1 \; = \; [q^1 \; u^2 \; u^3 \; ... \; u^n] \tag{137}$$

Now consider the product of A and Q_1.

$$\begin{aligned} AQ_1 &= (Aq^1 \; Au^2 \; Au^3 ... \; Au^n) \\ &= (\lambda_j q^1 \; Au^2 \; Au^3 ... \; Au^n) \end{aligned} \tag{138}$$

Now premultiply AQ$_1$ by ${Q_1}'$

$$Q_1{}' A Q_1 = \begin{pmatrix} \lambda_j & {q^1}' Au^2 & {q^1}' Au^3 & \cdots & {q^1}' Au^n \\ \lambda_j {u^2}' q^1 & {u^2}' Au^2 & {u^2}' Au^3 & \cdots & {u^2}' Au^n \\ \lambda_j {u^3}' q^1 & {u^3}' Au^2 & {u^3}' Au^3 & \cdots & {u^3}' Au^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_j {u^n}' q^1 & {u^n}' Au^2 & {u^n}' Au^3 & \cdots & {u^n}' Au^n \end{pmatrix} \tag{139}$$

The last n-1 elements of the first column are zero by orthogonality of q$_1$ and the vectors u. Similarly because A is symmetric and ${q^1}' Au^i$ is a scalar, the first row is zero except for the first element because

$$\begin{aligned} {q^1}' Au^i &= {u^i}' A q^1 \\ &= {u^i}' \lambda_j q^1 \\ &= \lambda_j {u^i}' q^1 \; = \; 0 \end{aligned} \tag{140}$$

Now rewrite ${Q_1}' A Q_1$ as

$$Q_1' \, AQ_1 = \begin{pmatrix} \lambda_j & 0 & 0 & \cdots & 0 \\ 0 & {u^2}' Au^2 & {u^2}' Au^3 & \cdots & {u^2}' Au^n \\ 0 & {u^3}' Au^2 & {u^3}' Au^3 & \cdots & {u^3}' Au^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & {u^n}' Au^2 & {u^n}' Au^3 & \cdots & {u^n}' Au^n \end{pmatrix} \tag{141}$$

$$= \begin{bmatrix} \lambda_j & \mathbf{0} \\ \mathbf{0} & \alpha_1 \end{bmatrix} = A_1$$

where α_1 in an (n-1)x(n-1) symmetric matrix. Because Q_1 is an orthogonal matrix its inverse is equal to its transpose so that A and A_1 are similar matrices (see theorem 3) and have the same characteristic roots. With the same characteristic roots (one of which is here denoted by λ) we obtain

$$|\, \lambda I_n \, - \, A \,| \; = \; |\, \lambda I_n \, - \, A_1 \,| \; = \; 0 \tag{142}$$

Now consider cases where $k_1 \geq 2$, that is the root λ_j is not distinct. Write the equation $|\, \lambda I_n - A_1 \; = \; 0 \,|$ in the following form.

$$\begin{aligned} |\, \lambda I_n \, - \, A_1 \,| &= \left| \begin{bmatrix} \lambda & 0 \\ 0 & \lambda I_{n-1} \end{bmatrix} - \begin{bmatrix} \lambda_j & 0 \\ 0 & \alpha_1 \end{bmatrix} \right| \\ &= \begin{vmatrix} \lambda - \lambda_j & 0 \\ 0 & \lambda I_{n-1} \, - \, \alpha_1 \end{vmatrix} \end{aligned} \tag{143}$$

We can compute this determinant by a cofactor expansion of the first row. We obtain

$$|\, \lambda I_n \, - \, A \,| \; = \; (\, \lambda \, - \, \lambda_j \,) \; |\, \lambda I_{n-1} \, - \, \alpha_1 \,| \; = \; 0 \tag{144}$$

Because the multiplicity of $\lambda_j \geq 2$, at least 2 elements of λ will be equal to λ_j. If a characteristic root that solves equation 144 is not equal to λ_j then $|\, \lambda I_{n-1} \, - \, \alpha_1 \,|$ must equal zero, i.e.,if $\lambda \neq \lambda_j$

$$|\, \lambda_j I_{n-1} \, - \, \alpha_1 \,| \; = \; 0 \tag{145}$$

If this determinant is zero then all minors of the matrix

$$\lambda_j I_n \, - \, A_1 \; = \; \begin{bmatrix} 0 & 0 \\ 0 & \lambda_j I_{n-1} \, - \, \alpha_1 \end{bmatrix} \tag{146}$$

of order n-1 will vanish. This means that the nullity of $\lambda_j I_n \, - \, A_1$ is at least two. This is true because the nullity of the submatrix $\lambda_j \mathrm{I}_{n-1}$ - α is at least one and it is a submatrix of A_1 - $\lambda_j \mathrm{I}_{n-1}$ which has a row and column of zeroes and thus already has a nullity of one. Because $rank(\lambda_j I_n \, - \, A) = rank(\lambda_j I_n \, - \, A_1)$, the nullity of $\lambda_j I_n \, - \, A$ is ≥ 2. This means that we can find another characteristic vector q^2, that is linearly independent of q^1 and is also orthogonal to q^1. If the multiplicity of $\lambda_j = 2$, we are finished.

Otherwise define the matrix Q_2 as

$$Q_2 \; = \; [q^1 \; q^2 \; u^3 \; ... \; u^n] \tag{147}$$

Now let A_2 be defined as

$$\begin{aligned} Q_2' \, A Q_2 \; &= \begin{pmatrix} \lambda_j & 0 & 0 & \cdots & 0 \\ 0 & \lambda_j & 0 & \cdots & 0 \\ 0 & u^{3'} Au^2 & u^{3'} Au^3 & \cdots & u^{3'} Au^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & u^{n'} Au^2 & u^{n'} Au^3 & \cdots & u^{n'} Au^n \end{pmatrix} \\ &= \begin{bmatrix} \lambda_j & 0 & 0 \\ 0 & \lambda_j & 0 \\ 0 & 0 & \alpha_2 \end{bmatrix} \end{aligned} \tag{148}$$

where α_2 in an (n-2)x(n-2) symmetric matrix. Because Q_2 is an orthogonal matrix its inverse is equal to its transpose so that A and A_2 are similar matrices and have the same characteristic roots, that is.

$$|\, \lambda I_n \, - \, A \,| \; = \; |\, \lambda I_n \, - \, A_2 \,| \; = \; 0 \tag{149}$$

Now consider cases where $k_1 \geq 2$, that is the root λ_j is not distinctand has multiplicity ≥ 3. Write the equation $|\, \lambda I_n \, - \, A_2 \; = \; 0 \,|$ in the following form.

$$\begin{aligned} |\, \lambda I_n \, - \, A_2 \,| \; &= \left| \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda I_{n-2} \end{bmatrix} - \begin{bmatrix} \lambda_j & 0 & 0 \\ 0 & \lambda_j & 0 \\ 0 & 0 & \alpha_2 \end{bmatrix} \right| \\ &= \begin{vmatrix} \lambda - \lambda_j & 0 & 0 \\ 0 & \lambda - \lambda_j & 0 \\ 0 & 0 & \lambda I_{n-2} \, - \, \alpha_2 \end{vmatrix} \end{aligned} \tag{150}$$

If we expand this by the first row and then again by the first row we obtain

$$|\, \lambda I_n \, - \, A \,| \; = \; (\lambda \, - \, \lambda_j \,)^2 \; |\, \lambda I_{n-2} \, - \, \alpha_2 \,| \; = \; 0 \tag{151}$$

Because the multiplicity of $\lambda_j \geq 3$, at least 3 elements of λ will be equal to λ_j. If a characteristic root that solves equation 151 is not equal to λ_j then $|\,\lambda I_{n-2} \,-\, \alpha_2\,|$ must equal zero, i.e.,if $\lambda \neq \lambda_j$

$$|\,\lambda_j I_{n-2} \,-\, \alpha_2\,| \,=\, 0 \tag{152}$$

If this determinant is zero then all minors of the matrix

$$\lambda_j I_n \,-\, A_2 \,=\, \begin{bmatrix} 0 & 0 & \mathbf{0} \\ 0 & 0 & \mathbf{0} \\ 0 & 0 & \lambda_j I_{n-1} \,-\, \alpha_2 \end{bmatrix} \tag{153}$$

of order n-2 will vanish. This means that the rank of matrix is less than n-2 and so its nullity is greater than or equal to 3. This means that we can find another characteristic vector q^3, that is linearly independent of q^1 and q^2 and is also orthogonal to q^1. If the multiplicity of λ_j = 3, we are finished. Otherwise, proceed as before. Continuing in this way we can find k_j orthonormal vectors.

Now we need to show that we cannot choose more than k_j such vectors. After choosing q^m, we would have

$$|\,\lambda I_n \,-\, A\,| \,=\, |\,\lambda I_n \,-\, A_{k_j}\,| = \begin{vmatrix} (\lambda - \lambda_j)\, I_{k_j} & 0 \\ 0 & \lambda\, I_{n-k_j} \,-\, \alpha_{k_j} \end{vmatrix} \tag{154}$$

If we expand this we obtain

$$|\,\lambda I_n \,-\, A\,| \,=\, (\lambda \,-\, \lambda_j)^{k_j} \, \left|\,\lambda I_{n-k_j} \,-\, \alpha_{k_j}\,\right| \,=\, 0 \tag{155}$$

It is evident that

$$|\,\lambda I_n \,-\, A\,| \,=\, (\lambda \,-\, \lambda_j\,)^{k_j} \, \left|\,\lambda I_{n-k_j} \,-\, \alpha_{k_j}\,\right| \,=\, 0 \tag{156}$$

implies

$$\lambda_j I_{n-k_j} \,-\, \alpha_{k_j}\,| \,\neq\, 0 \tag{157}$$

If this determinant were zero the multiplicity of λ_j would exceed k_j. Thus

$$rank\,(\lambda_j I \,-\, A) \,=\, n \,-\, k_j \tag{158}$$

$$\tag{159}$$

$$nullity\,(\lambda_j I \,-\, A) \,=\, k_j \tag{160}$$

And this implies that the vectors we have chosen form a basis for the null space of λ_j - A and any additional vectors would be linearly dependent.

□

Corollary 1. Let A be a matrix as in the above theorem. The multiplicity of the root λ_j is equal to the nullity of λ_jI - A.

Theorem 22. *Let S be a symmetric matrix of order n. Then the characteristic vectors of S can be chosen to be an orthonormal set, i.e., there exists a Q such that*

$$Q'SQ \,=\, \Lambda \quad \textit{where Q is orthonormal} \tag{161}$$

Proof. Let the distinct characteristic roots of S be λ_j, j = 1, 2,... s $\leq$ n. The multiplicity of λ_j is given by k_j and

$$\Sigma_{j=1}^{s} k_j \; = \; n \tag{162}$$

By the corollary above the multiplicity of λ_j, k_j, is equal to the nullity of λ_j - S. By theorem 21 there exist k_j orthonormal characteristic vectors corresponding to λ_j. By theorem 5 the characteristic vectors corresponding to distinct roots are linearly independent. Hence the matrix

$$Q \; = \; (q^1 \; q^2 \; ... \; q^n) \tag{163}$$

where the first k_1 columns are the characteristic vectors corresponding to λ_1, the next k_2 columns correspond to λ_2, and so on is an orthogonal matrix. Now define the following matrix Λ

$$\begin{aligned} \Lambda \; &= diag(\lambda_1 I_{k_1} \; \lambda_2 I_{k_2}) \\ &= \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix} \end{aligned} \tag{164}$$

Note that

$$SQ \; = \; Q\Lambda \tag{165}$$

because Λ is the matrix of characteristic roots associated with the characteristic vectors Q. Now premultiply both sides of 165 by $Q^{-1} = Q'$ to obtain

$$Q'SQ \; = \; \Lambda \tag{166}$$

□

Consider the implication of theorem 22. Given a matrix Σ, we can convert it to a diagonal matrix by premultiplying and postmultiplying by an orthonormal matrix Q. The matrix Σ is

$$\Sigma \; = \; \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1n} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^{n1} & \sigma^{n2} & \cdots & \sigma^{nn} \end{bmatrix} \tag{167}$$

And the matrix product is

$$Q' \, \Sigma Q \; = \; \begin{bmatrix} s^{11} & 0 & \cdots & 0 \\ 0 & s^{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & s^{nn} \end{bmatrix} \tag{168}$$

10. Idempotent matrices and characteristic roots

Definition 6. A matrix A is idempotent if it is square and AA = A.

Theorem 23. *Let A be a square idempotent matrix of order n. Then its characteristic roots are either zero or one.*

Proof. Consider the equation defining characteristic roots and multiply it by A as follows

$$\begin{aligned} Ax &= \lambda x \\ AAx &= \lambda Ax \\ &= \lambda^2 x \\ \Rightarrow\ Ax &= \lambda^2 x \qquad \text{because A is idempotent} \end{aligned} \tag{169}$$

Now multiply both sides of equation 169 by x'.

$$\begin{aligned} Ax &= \lambda^2 x \\ x'\,Ax &= \lambda^2 x'\,x \\ \Rightarrow\ x'\,x\lambda &= \lambda^2 x'\,x \\ \Rightarrow\ \lambda &= \lambda^2 \\ \Rightarrow\ \lambda &= 1\ or\ \lambda\ =\ 0 \end{aligned} \tag{170}$$

□

Theorem 24. *Let A be a square symmetric idempotent matrix of order n and rank r. Then the trace of A is equal to the rank of A, i.e., tr(A) = r(A).*

To prove this theorem we need a lemma.

Lemma 1. *The rank of a symmetric matrix S is equal to the number of non-zero characteristic roots.*

Proof of lemma 1 . Because S is symmetric, it can be diagonalized using an orthogonal matrix Q, i.e.,

$$Q'SQ\ =\ \Lambda \tag{171}$$

Because Q is orthogonal, it is non-singular. We can rearrange equation 171 to obtain the following expression for S using the fact that Q^{-1} = Q′ and ${Q'}^{-1}$ = Q.

$$\begin{aligned} Q'SQ &= \Lambda \\ QQ'SQ &= Q\Lambda \qquad \text{(premultiply by Q = } {Q'}^{-1}\text{)} \\ QQ'SQQ' &= Q\Lambda Q' \qquad \text{(postmultiply by Q}' = Q^{-1}\text{)} \\ \Rightarrow\ S &= Q\Lambda Q' \qquad \text{(QQ}' = \text{I)} \end{aligned} \tag{172}$$

Given that S = QΛQ′, the rank of S is equal to the rank of QΛQ′. The multiplication of a matrix by a non-singular matrix does not affect its rank so that the rank of S is equal to the rank of Λ. But the only non-zero elements in the diagonal matrix Λ are the non-zero characteristic roots. Thus the rank of the matrix Λ is equal to the number of non-zero characteristic roots. This result actually holds for all diagonalizable matrices, not just those that are symmetric..

□

Proof of theorem 24 . Consider now an idempotent matrix A. Use the orthogonal transformation to diagonalize it as follows

$$Q'\,AQ \;=\; \Lambda \;=\; \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{173}$$

All the elements of Λ will be zero or one because A is idempotent. For example the matrix might look like this

$$Q'\,AQ \;=\; \Lambda \;=\; \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{174}$$

Now the number of non-zero roots is just the sum of the ones. Furthermore the sum of the characteristic roots of a matrix is equal to the trace. Thus the trace of A is equal to its rank.

□

11. Simultaneous Diagonalization of Matrices

Theorem 25. *. Suppose that A and B are m x m symmetric matrices. Then there exists an orthogonal matrix P such that $P'AP$ and $P'BP$ are both diagonal if and only if A and B commute, that is, if and only if AB = BA.*

Proof. First suppose that such an orthogonal matrix does exist; that is, there is an orthogonal matrix P such as $P'AP = \Lambda_1$ and $P'BP = \Lambda_2$, where Λ_1 and Λ_2 are diagonal matrices. Given that P is orthogonal, by Theorem 11 we have $P' = P^{-1}$ and $P'^{-1} = P$. We can then write A in terms of P and Λ and B in terms of P and Λ as follows

$$\begin{aligned} \Lambda_1 &= P'\,A\,P \\ \Rightarrow\; \Lambda_1\,P^{-1} &= P'A \\ \Rightarrow\; P'^{-1}\,\Lambda_1\,P^{-1} &= A \\ \Rightarrow\; P\,\Lambda_1\,P' &= A \end{aligned} \tag{175}$$

for A, and

$$\begin{aligned} \Lambda_2 &= P'\,B\,P \\ \Rightarrow\; P'^{-1}\,\Lambda_2\,P^{-1} &= B \\ \Rightarrow\; P\,\Lambda_2\,P' &= B \end{aligned} \tag{176}$$

for B. Because Λ_1 and Λ_2 are diagonal matrices we have $\Lambda_1\,\Lambda_2 \;=\; \Lambda_2\,\Lambda_1$. Using this fact we can write AB as

$$\begin{aligned} AB &= P\Lambda_1 P' P \Lambda_2 P' \\ &= P\Lambda_1 \Lambda_2 P' \\ &= P\Lambda_2 \Lambda_1 P' \\ &= P\Lambda_2 P' P \Lambda_1 P' \\ &= BA \end{aligned} \tag{177}$$

and hence, A and B do commute.

Conversely, now assuming that AB = BA, we need to show that such an orthogonal matrix P does exist. Let $\mu_1, ..., \mu_h$ be the distinct values of the characteristic roots of A having multiplicities $r_1, ..., r_h$, respectively. Because A is symmetric there exists an orthogonal matrix Q satisfying

$$\begin{aligned} Q'AQ &= \Lambda_1 \;=\; diag\,(\mu_1 I_{r_1}, \mu_2 I_{r_2}, \ldots, \mu_h I_{r_h}) \\ &= \begin{pmatrix} \mu_1 I_{r_1} & 0 & 0 & \cdots & 0 \\ 0 & \mu_2 I_{r_2} & 0 & \cdots & 0 \\ 0 & 0 & \mu_3 I_{r_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mu_h I_{r_h} \end{pmatrix} \end{aligned} \tag{178}$$

If the multiplicity of each root is one, then we can write

$$Q'AQ \;=\; \Lambda_1 \;=\; \begin{bmatrix} \mu_1 & 0 & \cdots & 0 & 0 \\ 0 & \mu_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \mu_h \end{bmatrix} \tag{179}$$

Performing this same transformation on B and partitioning the resulting matrix in the same way that Q'AQ has been partitioned, we obtain

$$C \;=\; Q'BQ \;=\; \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1h} \\ C_{21} & C_{22} & \cdots & C_{2h} \\ \vdots & \vdots & & \vdots \\ C_{h1} & C_{h2} & \cdots & C_{hh} \end{bmatrix} \tag{180}$$

where C_{ij} is r_i x r_j. For example, C_{11} is square and will have dimension equal to the multiplicity of the first characteristic root of A. C_{12} will have the same number of rows as the multiplicity of the first characteristic root of A and number of columns equal to the multiplicity of the second characteristic root of A. Given that AB = BA, we can show that $\Lambda_1 C \;=\; C\Lambda_1$. To do this we write $\Lambda_1 C$, substitute for Λ_1 and C, simplify, make the substitution and reconstitute.

$$\begin{aligned} \Lambda_1 C &= (Q'AQ)\,(Q'BQ) && (definition) \\ &= Q'A(QQ')BQ && (regroup) \\ &= Q'ABQ && (QQ' = I) \\ &= Q'BAQ && (AB = BA) \\ &= Q'BQQ'AQ && (QQ' = I) \\ &= C\Lambda_1 && (definition) \end{aligned} \tag{181}$$

Equating the (i,j)th submatrix of λ_1C to the (i,j)th submatrix of Cλ_1 yields the identity $\mu_i\, C_{ij}$ = $\mu_j\, C_{ij}$. Because $\mu_i \neq \mu_j$ if i $\neq$ j, we must have C_{ij} = 0 if i $\neq$ j; that is C is not a densely populated matrix but rather is,

$$
\begin{aligned}
C \; &= diag(C_{11}\,,\; C_{22}\,,\; \ldots\,,\; C_{hh}\,) \\
&= \begin{bmatrix} C_{11} & 0 & \cdots & 0 & 0 \\ 0 & C_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & C_{hh} \end{bmatrix}
\end{aligned} \tag{182}
$$

Now because C is symmetric so also is C_{ii} for each i, and thus, we can find an $r_i \times r_i$ orthogonal matrix X_i (by theorem 22) satisfying

$$X'\, C_{ii}\, X_i \;=\; \Delta_i \tag{183}$$

where Δ_i is diagonal. Let P = QX, where X is the block diagonal matrix X = diag($X_1, X_2, \ldots . X_h$), that is

$$X \;=\; \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X_h \end{bmatrix} \tag{184}$$

Now write out P'P, simplify and substitute for X as follows

$$
\begin{aligned}
P'\,P \; &= X'\,Q'\,Q\,X \\
&= X'\,X \\
&= \begin{bmatrix} X_1' & 0 & 0 & \cdots & 0 \\ 0 & X_2' & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X_h' \end{bmatrix} \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X_h \end{bmatrix} \\
&= diag\,(X_1'\,X_1,\; X_2'\,X_2, \ldots, X_h'\,X_h) \\
&= diag\,(\,I_{r_1}\,,\; I_{r_2}, \ldots, I_{r_h}) \\
&= I_m
\end{aligned} \tag{185}
$$

Given that P'P is an identity matrix, P is orthogonal. Finally, the matrix $\Delta = diag(\Delta_1, \Delta_2, \ldots, \Delta_h)$ is diagonal and

$$
\begin{aligned}
P'AP \;=\; (X'Q')A(QX) \; &= X'(Q'AQ)X \\
&= X'\,\Lambda_1\,X \\
&= diag\,(X_1', X_2', \ldots, X_h')\,diag\,(\,\mu_1 I_{r_1},\; \mu_2 I_{r_2}, \ldots, \mu_h I_{r_h}), diag\,(X_1, X_2, \ldots, X_h) \\
&= diag\,(\mu_1\,X_1'\,X_1\,,\; \mu_2\,X_2'\,X_2\,,\; \ldots\,,\; \mu_{h\,X_h}\prime\,X_h\,) \\
&= diag\,(\,\mu_1\,I_{r_1}\,,\; \mu_2\,I_{r_2}\,, \ldots\,,\; \mu_{h\,I_{r_h}}\,) \\
&= \Lambda_1
\end{aligned} \tag{186}
$$

and

$$
\begin{aligned}
P' \, B \, P \; &= X' \, Q' \, B \, Q \, X \; = \; X' \, \Lambda_2 \, X \\
&= diag \, (X_1' \, , \; X_2' \, , \; \ldots \, , \; X_h' \;) \; diag \, (\, C_{11} \, , \; C_{22} \, , \ldots \, , \; C_{hh} \;) \; diag \, (X_1 \, , \; X_2 \, , \; \ldots \, , \; X_h \;) \\
&= diag \, (X_1' \, C_{11} \, X_1 \, , \; X_2' \, C_{22} \, X_2 \, , \; \ldots \, , \; X_h' \, C_{hh} \, X_h \,) \\
&= diag \, (\, \Delta_{11} \, , \; \Delta_{22} \, , \ldots \, , \; \Delta_{hh} \,) \\
&= \Delta
\end{aligned} \tag{187}
$$

This then completes the proof.

□

References

[1] Dhrymes, P.J. *Mathematics for Econometrics - 3^{rd} Edition.* New York: Springer-Verlag, 2000
[2] Gantmacher, F.R. *The Theory of Matrices, Vol 1).* New York: Chelsea Publishing Company, 1977
[3] Hadley, G. *Linear Algebra.* Reading: Addison Wesley Publishing Company, 1961
[4] Horn, R.A. and C.R. Johnson. *Matrix Analysis.* Cambridge: Cambridge University Press, 1985
[5] Schott, J. R. *Matrix Analysis for Statistics*. New York: Wiley, 1997
[6] Searle, S. R. *Matrix Algebra Useful for Statistics.* New York: Wiley, 1982
[7] Sydsaeter, K. *Topics in Mathematical Analysis for Economists.* London: Academic Press, 1981