

Math 642:550 — Summer 2002  
MTTh 6:15–8:45 PM Hill 525  
Prof. Bumby

## The Perron-Frobenius Theorem.

**Introduction.** Square matrices whose entries are all nonnegative have special properties. This was mentioned briefly in Section 5.3 with a sketch of the proof when all entries of the matrix are positive appearing as item 5K. The result was first proved by Perron in the case when all entries are positive. Frobenius then filled in all the details to identify the nature of all exceptions. We will follow the history of the subject by dealing with positive matrices before examining the modifications needed for general nonnegative matrices.

The theorem asserts that there is a eigenvector of such a matrix, all of whose entries are nonnegative, and then goes on to give further properties of this eigenvector and its eigenvalue.

Markov matrices  $M$  are a special case, and we have seen that a probabilistic interpretation of the solution of  $Mv = v$  based on a claim that all entries of  $v$  are nonnegative. In keeping with this probabilistic interpretation, it is customary to scale the nonnegative eigenvector so that the sum of its entries is 1, rather than use the *Euclidean* scaling that has the sum of the *squares* of the entries equal to 1 that is so useful in the theory of symmetric matrices.

These notes include material from F. R. Gantmacher, *The Theory of Matrices*, Chelsea, 1990, Chapter XIII and Roger A. Horn & Charles R. Johnson, *Matrix Analysis*, Cambridge, 1985, Chapter 8. However, all expositions aimed at mathematicians emphasize an efficient path to a proof. This may not be the best approach for those more interested in applying the results. The subject also has features that relate this topic to other areas of linear algebra and suggest methods to approximate the positive eigenvector and its eigenvalue. These features will also be discussed.

We work in a fixed space, which can be taken to be  $\mathbb{R}^n$ , so that all vectors have  $n$  components and all matrices are  $n$  by  $n$ . Since we are concerned with eigenvalues and eigenvectors, the matrix must be considered as mapping this space to **itself**.

It is only the eigenvalue with a positive eigenvector that is distinguished in this theory. It will be called the **dominant** eigenvalue, since the other eigenvalues will be shown to be of **smaller absolute value**. In particular, the remaining eigenvalues need not be real.

**The Role of nonnegativity.** Write  $v \geq w$  to denote the property that **no entry** of the vector  $v$  is smaller than the corresponding entry of  $w$ . If every entry of a matrix  $A$  is nonnegative, then  $v \geq w$  implies  $Av \geq Aw$  since the difference of  $i^{\text{th}}$  entries is

$$\sum_{j=1}^n a_{ij}(v_j - w_j)$$

and all parts of this expression are assumed nonnegative. Furthermore, the same expression shows that, if every entry of  $A$  is positive, and *some* entry of  $v$  is strictly greater than the corresponding entry of  $w$ , then **every entry** of  $Av$  is strictly greater than the corresponding entry of  $Aw$ .

This leads to the following observation that we will need in the proof, so we give it a name.

**Lemma 1.** If all entries of  $A$  are positive and  $v$  is an eigenvector of  $A$  with  $v \geq 0$ , that all entries of  $v$  are strictly positive.

*Proof.* Since all entries of  $v$  are nonnegative, the same is true of  $Av$ . Furthermore, all entries of  $A$  are positive, so the entries of  $Av$  are all positive. However,  $v$  is an eigenvector, so  $Av$  is a scalar multiple of  $v$  which requires it to have zero entries in the same locations as  $v$ . Hence, none of the entries of  $v$  can be zero.

Since eigenvectors of  $A$  are also eigenvectors of all powers of  $A$ , if some power of  $A$  has all entries positive, any eigenvector of  $A$  with nonnegative entries must have strictly positive entries.

**Uniqueness of positive eigenvectors.** Many important results are organized around an **existence and uniqueness theorem**. The proofs of such results are split into two parts: in the existence part, one proves that **at least one** example exists; so in this context, the uniqueness proof need only show that **at most one** example exists. Frequently, **uniqueness, in this sense, is easier to prove than existence**, so it is common to give this part of the proof first. The use of these words outside of mathematics suggests that uniqueness is only a meaningful question after existence has been settled, but the mathematical interpretation makes it reasonable to reverse the order of proofs. In particular, if you know that there is at most one solution to the problem, it becomes reasonable to seek a constructive method to find it that can be used in the existence proof.

The uniqueness theorem in these notes will be rather weak because it has a strong hypothesis, but that hypothesis will be justified by the existence theorem.

**Uniqueness Theorem.** If a matrix  $A$  with nonnegative entries (but not the zero matrix) has a positive **row eigenvector**  $u$  with eigenvalue  $\lambda$ , then every nonnegative eigenvector  $v$  has eigenvalue  $\lambda$ .

*Proof.* First, note that  $uv > 0$ . Then, if  $v$  has eigenvalue  $\mu$ , we have

$$\lambda uv = (uA)v = u(Av) = \mu uv,$$

and the desired conclusion follows.

**Corollary.** If  $M$  is a Markov matrix, then a nonnegative eigenvector has eigenvalue 1.

*Proof.* The definition gives  $u = [1, \dots, 1]$  and  $\lambda = 1$ .

**Corollary.** Under the assumptions of the theorem, if  $A$  has all positive entries, then the only eigenvectors with eigenvalue  $\lambda$  are the multiples of  $v$ , i.e.  $v$  has geometric multiplicity 1..

*Proof.* Lemma 1 shows that  $v$  is positive. If  $w$  were any other eigenvector with eigenvalue  $\lambda$ , then there would be a vector of the form  $v + \alpha w$  with all entries nonnegative and at least one zero entry. However, this also contradicts lemma 1.

**Absolute values of eigenvectors.** Suppose that we have an eigenvector  $v$ , not necessarily positive, of a nonnegative matrix  $A$  with  $Av = \lambda v$ . If the  $i^{\text{th}}$  entry of  $v$  is denoted  $v_i$ , let  $v^+$  denote the vector whose  $i^{\text{th}}$  entry is  $|v_i|$ . Let  $w_i$  denote the  $i^{\text{th}}$  entry of  $Av^+$ . Then, the triangle inequality

gives

$$w_i = \sum_{j=1}^n a_{ij} v_j^+ = \sum_{j=1}^n |a_{ij}| |v_j| \geq \left| \sum_{j=1}^n a_{ij} v_j \right| = |\lambda v_i| = |\lambda| v_i^+$$

Thus, any number  $\lambda$  that can be an eigenvalue leads to a nonnegative vector  $v$  such that  $Av \geq |\lambda| v$ . Thus, interest in eigenvalues leads to interest in the set of all positive numbers  $c$  for which there is a nonnegative vector  $v$  with  $Av \geq cv$ . Since the same  $v$  also gives this relation for all smaller  $c$ , this set is an interval. This interval is bounded since, if the largest  $v_i$  is  $v_k$ , then the  $k^{\text{th}}$  entry of  $Mv$  is bounded by

$$\sum_{j=1}^n a_{kj} v_j \leq \sum_{j=1}^n a_{kj} v_k \leq \sup_i \left( \sum_{j=1}^n a_{ij} \right) v_k.$$

In particular,  $c$  can be no larger than an easily computed **matrix norm** of  $A$ .

The usual proof of existence relies on the existence of a least upper bound for bounded sets of real numbers to produce a right-hand endpoint of the interval of such  $c$  and then to show that this endpoint must be an eigenvalue by extracting a convergent subsequence of the sequence of vectors  $v$  associated with these  $c$ . Although technical, this analysis is a familiar tool in current mathematics.

For positive matrices, the eigenvalues with eigenvectors that are not strictly positive must be strictly smaller than the eigenvalue found in this way.

Since this applies equally well to  $A^T$ , this gives the row eigenvector of  $A$  used in the proof of the uniqueness theorem.

**Approximate eigenvectors.** Suppose that we take a value of  $c$  close to, but smaller than, the dominant eigenvalue. There are positive vectors  $v$  with  $Av \geq cv$ .

As in the construction of the characteristic polynomial, rewrite  $Av \geq cv$  as  $(A - cI)v \geq 0$ . This is a system of inequalities with positive coefficients **except on the diagonal**. We describe an inductive construction on the matrix  $(A - cI)$  that provides a test for the system to have a nonnegative solution.

If the first diagonal entry is negative, pivoting on this position adds a positive multiple of the first row to all other rows. If the expressions represented by the rows of the original matrix are all nonnegative, the same will be true of the result of this pivot step. The rows other than the first row of this new matrix represent a system of inequalities not containing the first variable. The induction shifts attention to that system.

If the only entry of a 1 by 1 matrix is negative, this represents an inequality that asserts that a negative multiple of a nonnegative variable is nonnegative. Such an inequality has no solution, so any system of inequalities with the property that a solution to that system would lead to a solution of this inequality must also have no solutions. Thus, if every pivot of  $(A - cI)$  is negative, the only nonnegative vector  $v$  with  $Av \geq cv$  is the zero vector. This implies that  $c$  is greater than the dominant eigenvalue of  $A$ .

On the other hand, suppose that we meet a nonnegative pivot in the  $k^{\text{th}}$  row. This will allow us to produce a solution of the original system of inequalities: we do three things (although only the action on the  $k^{\text{th}}$  variable need actually be **done**, the others may be vacuous). First, the variables after the  $k^{\text{th}}$  variable are set equal to zero. All inequalities of the **original system** after the  $k^{\text{th}}$  row now assign only nonnegative coefficients to the remaining variables, so they are automatically satisfied if these variables are nonnegative. This restricts attention to the  $k$  by  $k$  upper left submatrix. Next, require expressions represented by the first  $k - 1$  rows of the original system to be equal to zero. This means that all the pivot steps that have been done **preserve** the values represented by each of the  $k$  rows for any choice of the variables. When we enter the back-substitution phase, these equations will be used to determine the first  $k - 1$  variable in terms of the  $k^{\text{th}}$

variable, and they give nonnegative values to all of these variables. There remains one row of the reduced system that asserts that a nonnegative multiple of the  $k^{\text{th}}$  variable is nonnegative. This can be satisfied by requiring this variable to be positive (since the system is homogeneous, it can be set equal to 1 without loss of generality). Now, do back substitution to find values of all variables that will make the first  $k - 1$  rows represent zero and row  $k$  positive — in both the original and the reduced systems.

If  $c$  is the largest value admitting a nonnegative vector  $v$  with  $Av \geq cv$ , then all pivots **except the last** will be negative, and the last pivot will be zero. In this case, we are solving **equations** and produce a nonnegative solution to  $Av = cv$ , which reveals the desired eigenvalue and eigenvector. The fact that only the last pivot is zero in this case can be used to prove that  $c$  has **algebraic** multiplicity 1.

If  $n$  is not too large, this allows us to decide how any given  $c$  compares to the dominant eigenvalue.

Our conclusion is that  $c$  is greater than the dominant eigenvalue if and only if all pivots are negative. As with a comparable characterization of positive definite matrices, this is equivalent to the sign of the determinant of the  $k$  by  $k$  upper left matrix being  $(-1)^k$ . If we are only interested in **testing**  $c$ , the pivot steps need not be done, any method may be used to compute this sequence of determinants.

**General nonnegative entries.** If not all entries of  $A$  are positive, there are other possibilities and these conclusions need to be weakened. Of course, if  $A^k$  is positive for some  $k$ , the full strength of the theorem applies to  $A^k$  whose eigenvalues are the  $k^{\text{th}}$  powers of the eigenvalues of  $A$ , so the results can be made to apply to  $A$ .

Instead of giving a detailed catalogue of modifications to get the theorem of Frobenius, we give some key examples.

The first collection of exceptions are those for which the matrix has an invariant coordinate subspace. For example,  $a_{1j}$  could be zero for all  $j > 2$ . A simple example is

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

In this case, if  $v$  is a vector with first component zero, so is  $Av$ . This example shows that this case allows generalized eigenvectors. However, a slight change in the diagonal entries of the matrix could restore a dominant eigenvalue. Such matrices are called **reducible**.

The second type of exception was illustrate by “Bernadelli’s beetle” in Strang’s exercise 5.3.4. Here,  $A$  itself is not reducible, but some  $A^k$  is. If this happens for some  $k$ , it also happens for all multiples of  $k$ , so there will be least  $k$  for which  $A^k$  is reducible. In this case,  $A^k$  is not just reducible, but in block diagonal form. Each block of  $A^k$  is an independent nonnegative matrix. Since these blocks are of smaller size than the original, we may suppose, by induction on the size of the matrix, that its eigenvalues and eigenvectors have been found. let  $\mu$  is an eigenvector and  $v$  is an eigenvector in one of the blocks, extended to be zero on the other blocks. Introduce  $\lambda$  to be any solution of  $\lambda^k = \mu$  and form

$$\sum_{j=0}^{k-1} \lambda^{-j} A^j v.$$

It is easy to verify that this is an eigenvector of eigenvalue  $\lambda$ . Hence, any such  $\lambda$  is an eigenvalue, and the nonzero eigenvalues produced from each block need to be the same (since  $A$  was assumed irreducible). Eigenvalues that are zero may be added to blocks arbitrarily. An example is

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 3 & 3 & 0 \end{bmatrix}$$

**Application to Population.** A **Leslie matrix** is a matrix  $A$  in which  $a_{ij} = 0$  unless  $i = 1$  or  $i = j + 1$ . This is used to model growth of a population represented by a vector of the population count by age. The matrix aims to describe the transition of this information from one year to the next. Here,  $a_{i(i+1)} = s_i$ , the probability that an individual of age  $i$  survives for a year (and becomes exactly one year older); and  $a_{1j}$  represents the average number of births where the parent has age  $j$ . This is a nonnegative matrix and some power can be shown to have all entries positive (as long as all the  $s_i > 0$ ). Thus, such a matrix has a positive eigenvector  $x = \langle x - i \rangle$ . The special form of the matrix shows that

$$s_i x_i = \lambda x_{i+1} \quad \text{and} \quad \sum_j a_{1j} x_j = \lambda x_1.$$

The case  $n = 4$  illustrates how this information can be used. The general case requires only a suitably general notation. The equations can be arranged as

$$\begin{aligned} x_3 &= \frac{\lambda}{s_3} x_4 \\ x_2 &= \frac{\lambda}{s_2} x_3 = \frac{\lambda}{s_2} \frac{\lambda}{s_3} x_4 \\ x_1 &= \frac{\lambda}{s_1} x_2 = \frac{\lambda}{s_1} \frac{\lambda}{s_2} \frac{\lambda}{s_3} x_4 \\ \lambda x_1 &= b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4. \end{aligned}$$

If the expressions in terms of  $x_4$  are substituted into the last equation, we get an equation that simplifies to

$$1 = b_1 \lambda^{-1} + b_2 s_1 \lambda^{-2} + b_3 s_1 s_2 \lambda^{-3} + b_4 s_1 s_2 s_3 \lambda^{-4}.$$

The quantity on the right is a decreasing function of  $\lambda$  that runs through all positive values as  $\lambda$  runs through all positive values. Since any eigenvalue must satisfy this polynomial, this characterizes the positive eigenvalue uniquely. In particular, we have  $\lambda < 1$  if and only if

$$b_1 + b_2 s_1 + b_3 s_1 s_2 + b_4 s_1 s_2 s_3 < 1.$$

The left side of this inequality may be interpreted as the expected number of offspring **over the lifetime of a newborn individual**. While this agrees with intuition, it should be noted that a **highly structured** average is used that depends on all  $b_j$  and all  $s_i$ . Discussions of population growth in general articles are likely to present an informal argument using a simpler, but less accurate, measure of the ability of the population to reproduce itself.

## Exercises.

Let

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 5 & 11 & 1 \\ 2 & 7 & 1 \end{bmatrix}$$

1. Verify that  $A^2$  has all entries positive. (This will show that Perron's theorem applies to  $A^2$ , determining the structure of eigenvalues and eigenvectors of  $A$ , and allowing computations to use  $A$ ).
2. Find a vector  $v$  with  $Av > 12v$  which shows that  $A$  must have an eigenvalue greater than 12. (Use only the methods described here — do not determine the characteristic polynomial.)
3. Show that no eigenvalue of  $A$  is as large as 13. (The same conditions apply as in exercise 2.)
4. Let  $v_1 = [1, 9, 6]^T$  and consider a sequence of vectors  $v_k$  with  $v_{i+1} = Av_i$ . Define  $c_i$  to be the largest quantity so that  $v_{i+1} \geq c_i v_i$ . Find  $v_1, v_2, v_3, v_4$  and use these to find  $c_1, c_2, c_3$ . Each of these is a lower bound on the dominant eigenvalue, so what is the largest lower bound that you now have on that quantity.