

Lecture: Basic Matrix Results (2 of 3)

*Lecturer: Michael Mahoney**Scribe: Michael Mahoney*

Warning: these notes are still very rough. They provide more details on what we discussed in class, but there may still be some errors, incomplete/imprecise statements, etc. in them.

3 Review and overview

Last time, we considered symmetric matrices, and we showed that if M is an $n \times n$ real-valued matrix, then the following hold.

- There are n eigenvalues, counting multiplicity, that are all real.
- The eigenvectors corresponding to different eigenvalues are orthogonal.
- Given k orthogonal eigenvectors, we can construct one more that is orthogonal to those k , and thus we can iterate this process to get a full set of n orthogonal eigenvectors
- This spectral theorem leads to a variational characterization of eigenvalues/eigenvectors and other useful characterizations.

These results say that symmetric matrices have several “nice” properties, and we will see that spectral methods will use these extensively.

Today, we will consider a different class of matrices and establish a different type of “niceness” result, which will also be used extensively by spectral methods. In particular, we want to say something about how eigenvectors, and in particular the extremal eigenvectors, e.g., the largest one or few or the smallest one or few “look like.” The reason is that spectral methods—both vanilla and non-vanilla variants—will rely crucially on this; thus, understanding when and why this is true will be helpful to see how spectral methods sit with respect to other types of methods, to understand when they can be generalized, or not, and so on.

The class of matrices we will consider are *positive matrices* as well as related *non-negative matrices*. By positive/non-negative, we mean that this holds element-wise. Matrices of this form could be, e.g., the symmetric adjacency matrix of an undirected graph, but they could also be the non-symmetric adjacency matrix of a directed graph. (In the latter case, of course, it is not a symmetric matrix, and so the results of the last class don’t apply directly.) In addition, the undirected/directed graphs could be weighted, assuming in both cases that weights are non-negative. In addition, it could apply more generally to any positive/non-negative matrix (although, in fact, we will be able to take a positive/non-negative matrix and interpret it as the adjacency matrix of a graph). The main theory that is used to make statements in this context and that we will discuss today and next time is something called Perron-Frobenius theory.

3.1 Some initial examples

Perron-Frobenius theory deals with positive/non-negative vectors and matrices, i.e., vectors and matrices that are entry-wise positive/nonnegative. Before proceeding with the main results of Perron-Frobenius theory, let us see a few examples of why it might be of interest and when it doesn't hold.

Example. Non-symmetric and not non-negative matrix. Let's start with the following matrix, which is neither positive/non-negative nor symmetric.

$$A = \begin{pmatrix} 0 & -1 \\ 2 & 3 \end{pmatrix}.$$

The characteristic polynomial of this matrix is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & -1 \\ 2 & 3 - \lambda \end{vmatrix} \\ &= -\lambda(3 - \lambda) + 2 \\ &= \lambda^2 - 3\lambda + 2 \\ &= (\lambda - 1)(\lambda - 2), \end{aligned}$$

from which it follows that the eigenvalues are 1 and 2. Plugging in $\lambda = 1$, we get $x_1 + x_2 = 0$, and so the eigenvector corresponding to $\lambda = 1$ is

$$x_{\lambda=1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Plugging in $\lambda = 2$, we get $2x_1 + x_2 = 0$, and so the eigenvector corresponding to $\lambda = 2$ is

$$x_{\lambda=2} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

So, this matrix has two eigenvalues and two eigenvectors, but they are not orthogonal, which is ok, since A is not symmetric.

Example. Defective matrix. Consider the following matrix, which is an example of a “defective” matrix.

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The characteristic polynomial is

$$\det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2,$$

and so 1 is a double root. If we plug this in, then we get the system of equations

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

meaning that $x_2 = 0$ and x_1 is arbitrary. (BTW, note that the matrix that appears in that system of equations is a nilpotent matrix. See below. From the last class, this has a value of the Rayleigh

quotient that is not in the closed interval defined by the min to max eigenvalue.) Thus, there is only one linearly independent eigenvector corresponding to the double eigenvalue $\lambda = 1$ and it is

$$x_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Example. Nilpotent matrix. Consider the following matrix,

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The only eigenvalue of this equals zero. The eigenvector is the same as in the above example. But this matrix has the property that if you raise it to some finite power then it equals the all-zeros matrix.

Example. Identity. The problem above with having only one linearly independent eigenvector is *not* due to the multiplicity in eigenvalues. For example, consider the following identity matrix,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

which has characteristic polynomial $\lambda^2 - 1 = 0$, and so which has $\lambda = 1$ as a repeated root. Although it has a repeated root, it has two linearly independent eigenvectors. For example,

$$x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

or, alternatively,

$$x_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad x_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

This distinction as to whether there are multiple eigenvectors associated with a degenerate eigenvalue is an important distinction, and so we introduce the following definitions.

Definition 1 *Given a matrix A , for an eigenvalue λ_i*

- *it's algebraic multiplicity, denoted $\mu_A(\lambda_i)$, is the multiplicity of λ as a root of the characteristic polynomial; and*
- *it's geometric multiplicity, denoted $\gamma_A(\lambda_i)$ is the maximum number of linearly independent eigenvectors associated with it.*

Here are some facts (and terminology) concerning the relationship between the algebraic multiplicity and the geometric multiplicity of an eigenvalue.

- $1 \leq \gamma_A(\lambda_i) \leq \mu_A(\lambda_i)$.
- If $\mu_A(\lambda_i) = 1$, then λ_i is a simple eigenvalue.
- If $\gamma_A(\lambda_i) = \mu_A(\lambda_i)$, then λ_i is a semi-simple eigenvalue.
- If $\gamma_A(\lambda_i) < \mu_A(\lambda_i)$, for some i , then the matrix A is defective. Defective matrices are more complicated since you need things like Jordan forms, and so they are messier.
- If $\sum_i \gamma_A(\lambda_i) = n$, then A has n linearly independent eigenvectors. In this case, A is diagonalizable. I.e., we can write $AQ = Q\Lambda$, and so $Q^{-1}AQ = \Lambda$. And conversely.

3.2 Basic ideas behind Perron-Frobenius theory

The basic idea of Perron-Frobenius theory is that if you have a matrix A with all positive entries (think of it as the adjacency matrix of a general, i.e., possibly directed, graph) then it is “nice” in several ways:

- there is one simple real eigenvalue of A that has magnitude larger than all other eigenvalues;
- the eigenvector associated with this eigenvalue has all positive entries;
- if you increase/decrease the magnitude of the entries of A , then that maximum eigenvalue increases/decreases; and
- a few other related properties.

These results generalize to non-negative matrices (and slightly more generally, but that is of less interest in general). There are a few gotchas that you have to watch out for, and those typically have an intuitive meaning. So, it will be important to understand not only how to establish the above statements, but also what the gotchas mean and how to avoid them.

These are quite strong claims, and they are certainly false in general, even for non-negative matrices, without those additional assumptions. About that, note that every nonnegative matrix is the limit of positive matrices, and so there exists an eigenvector with nonnegative components. Clearly, the corresponding eigenvalue is nonnegative and greater or equal in absolute value. Consider the following examples.

Example. Symmetric matrix. Consider the following matrix.

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This is a non-negative matrix, and there is an eigenvalue equal to 1. However, there exist other eigenvalues of the same absolute value (and not strictly less) as this maximal one. The eigenvalues are -1 and 1 , both of which have absolute value 1.

Example. Non-symmetric matrix. Consider the following matrix.

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This is a matrix in which the maximum eigenvalue is not simple. The only root of the characteristic polynomial is 0, and the corresponding eigenvector, i.e., $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, is not strictly positive.

These two counter examples contain the basic ideas underlying the two main gotchas that must be dealt with when generalizing Perron-Frobenius theory to only non-negative matrices.

(As an aside, it is worth wondering what is unusual about that latter matrix and how it can be generalized. One is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and there are others. These simple examples might seem trivial, but they contain several key ideas we will see later.)

One point of these examples is that the requirement that the entries of the matrix A be strictly positive is important for Perron-Frobenius theory to hold. If instead we only have non-negativity, we need further assumption on A which we will see below (and in the special case of matrices associated with graphs, the reducibility property of the matrix is equivalent to the connectedness of the graph).

3.3 Reducibility and types of connectedness

We get a non-trivial generalization of Perron-Frobenius theory from all-positive matrices to non-negative matrices, if we work with the class of irreducible matrices. (We will get an even cleaner statement if we work with the class of irreducible aperiodic matrices. We will start with the former first, and then we will bet to the latter.)

We start with the following definition, which applied to an $n \times n$ matrix A . For those readers familiar with Markov chains and related topics, there is an obvious interpretation we will get to, but for now we just provide the linear algebraic definition.

Definition 2 A matrix $A \in \mathbb{R}^{n \times n}$ is reducible if there exist a permutation matrix P such that

$$C = PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

with $A_{11} \in \mathbb{R}^{r \times r}$ and $A_{22} \in \mathbb{R}^{(n-r) \times (n-r)}$, where $0 < r < n$. (Note that the off-diagonal matrices, 0 and A_{12} , will in general be rectangular.) A matrix $A \in \mathbb{R}^{n \times n}$ is irreducible if it is not reducible.

As an aside, here is another definition that you may come across and that we may point to later.

Definition 3 A nonnegative matrix $A \in \mathbb{R}^{n \times n}$ is irreducible if $\forall i, j \in [n]^2, \exists t \in \mathbb{N} : A_{ij}^t > 0$. And it is primitive if $\exists t \in \mathbb{N}, \forall i, j \in [n]^2 : A_{ij}^t > 0$.

This is less intuitive, but I'm mentioning it since these are algebraic and linear algebraic ideas, and we haven't yet connected it with random walks. But later we will understand this in terms of things like lazy random walks (which is more intuitive for most people than the gcd definition of aperiodicity/primitiveness).

Fact: If A , a non-negative square matrix, is nilpotent (i.e., s.t. $A^k = 0$, for some $k \in \mathbb{Z}^+$, then it is reducible. *Proof:* By contradiction, suppose A is irreducible, and nilpotent. Let k be the smallest k such that $A^k = 0$. Then we know $A^{k-1} \neq 0$. Suppose $A_{ij}^{k-1} > 0$ for some i, j , since A irreducible, we now there exist $t \geq 1$ such that $A_{ji}^t > 0$. Note all powers of A are non-negative, then $A_{ii}^{k-1+t} = A_{i, \cdot}^{k-1} A_{\cdot, i}^t \geq A_{ij}^{k-1} A_{ji}^t > 0$ which gives a contradiction, since we have $A^k = 0 \Rightarrow A^{k'} = 0 \quad \forall k' \geq k$, while $k-1+t \geq k$, but $A^{k-1+t} \neq 0$.

◇

We start with a lemma that, when viewed the right way, i.e., in a way that is formal but not intuitive, is trivial to prove.

Lemma 1 *Let $A \in \mathbb{R}^{n \times n}$ be a non-negative square matrix. If A is primitive, then A is reducible.*

Proof: $\exists \forall \rightarrow \forall \exists$

◇

It can be shown that the converse is false. But we can establish a sort of converse in the following lemma. (It is a sort of converse since A and $I + A$ are related, and in particular in our applications to spectral graph theory the latter will essentially have an interpretation in terms of a lazy random walk associated with the former.)

Lemma 2 *Let $A \in \mathbb{R}^{n \times n}$ be a non-negative square matrix. If A is irreducible, then $I + A$ is primitive.*

Proof: Write out the binomial expansion

$$(I + A)^n = \sum_{k=0}^n \binom{n}{k} A^k.$$

This has all positive entries since A is irreducible, i.e., it eventually has all positive entries if k is large enough.

◇

Note that a positive matrix may be viewed as the adjacency matrix of a weighted complete graph.

Let's be more precise about directed and undirected graphs.

Definition 4 *A directed graph $G(A)$ associated with an $n \times n$ nonnegative matrix A consists of n nodes/vertices $P_1, \dots, P_n$, where an edge leads from P_i to P_j iff $A_{ij} \neq 0$.*

Since directed graphs are directed, the connectivity properties are a little more subtle than for undirected graphs. Here, we need the following. We will probably at least mention other variants later.

Definition 5 *A directed graph G is strongly connected if $\forall$ ordered pairs (P_i, P_j) of vertices of G , $\exists$ a path, i.e., a sequence of edges, $(P_i, P_{l_1}), (P_{l_1}, P_{l_2}), \dots, (P_{l_{r-1}}, P_j)$, which leads from P_i to P_j . The length of the path is r .*

Fact: The graph $G(A^k)$ of a nonnegative matrix A consists of all paths of $G(A)$ of length k (i.e. there is an edge from i to j in $G(A^k)$ iff there is a path of length k from i to j in G).

Keep this fact in mind since different variants of spectral methods involve weighting paths of different lengths in different ways.

Here is a theorem that connects the linear algebraic idea of irreducibility with the graph theoretic idea of connectedness. Like many things that tie together notions from two different areas, it can seem trivial when it is presented in such a way that it looks obvious; but it really is connecting two quite different ideas. We will see more of this later.

Theorem 1 *An $n \times n$ matrix A is irreducible iff the corresponding directed graph $G(A)$ is strongly connected.*

Proof: Let A be an irreducible matrix. Assume, for contradiction, that $G(A)$ is *not* strongly connected. Then, there exists an ordered pair of nodes, call them (P_i, P_j) , s.t. there does not exist a connection from P_i to P_j . In this case, let S_1 be the set of nodes connected to P_i , and let S_2 be the remainder of the nodes. Note that there is no connection between any nodes $P_\ell \in S_2$ and any node $P_q \in S_1$, since otherwise we could have $P_\ell \in S_1$. And note that both sets are nonempty, since $P_j \in S_1$ and $P_i \in S_2$. Let $r = |S_1|$ and $n - r = |S_2|$. Consider a permutation transformation $C = PAP^T$ that reorders the nodes of $G(A)$ such that

$$\begin{cases} P_1, P_2, \dots, P_r \in S_1 \\ P_{r+1}, P_{r+2}, \dots, P_n \in S_2 \end{cases}$$

That is

$$C_{k\ell} = 0 \quad \forall \begin{cases} k = r+1, r+2, \dots, n \\ \ell = 1, 2, \dots, r. \end{cases}$$

But this is a contradiction, since A is irreducible.

Conversely, assume that $G(A)$ is strongly connected, and assume for contradiction that A is not irreducible. Reverse the order of the above argument, and we arrive at the conclusion that $G(A)$ is not strongly connected, which is a contradiction.

◇

We conclude by noting that, informally, there are two types of irreducibility. To see this, recall that in the definition of reducibility/irreducibility, we have the following matrix:

$$C = PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}.$$

In one type, $A_{12} \neq 0$: in this case, we can go from the first set to the second set and get stuck in some sort of sink. (We haven't made that precise, in terms of random walk interpretations, but there is some sort of interaction between the two groups.) In the other type, $A_{12} = 0$: in this case, there are two parts that don't talk with each other, and so essentially there are two separate graphs/matrices.

3.4 Basics of Perron-Frobenius theory

Let's start with the following definition. (Note here that we are using subscripts to refer to elements of a vector, which is inconsistent with what we did in the last class.)

Definition 6 A vector $x \in \mathbb{R}^n$ is positive (resp, non-negative) if all of the entries of the vector are positive (resp, non-negative), i.e., if $x_i > 0$ for all $i \in [n]$ (resp if $x_i \geq 0$ for all $i \in [n]$).

A similar definition holds for $m \times n$ matrices. Note that this is *not* the same as SPD/SPSD matrices.

Let's also provide the following definition.

Definition 7 Let $\lambda_1, \dots, \lambda_n$ be the (real or complex) eigenvalues of a matrix $A \in \mathbb{C}^{n \times n}$. Then the spectral radius $\rho_A = \rho(A) = \max_i (|\lambda_i|)$

Here is a basic statement of the Perron-Frobenius theorem.

Theorem 2 (Perron-Frobenius) *Let $A \in \mathbb{R}^{n \times n}$ be an irreducible non-negative matrix. Then,*

1. *A has a positive real eigenvalue equal to its spectral radius.*
2. *That eigenvalue ρ_A has algebraic and geometric multiplicity equal to one.*
3. *The one eigenvector x associated with the eigenvalue ρ_A has all positive entries.*
4. *ρ_A increases when any entry of A increases.*
5. *There is no other non-negative eigenvector of A different than x .*
6. *If, in addition, A is primitive, then each other eigenvalue λ of A satisfies $|\lambda| < \rho_A$.*

Before giving the proof, which we will do next class, let's first start with some ideas that will suggest how to do the proof.

Let $P = (I + A)^n$. Since P is positive, it is true that for every non-negative and non-null vector v , that we have that $Pv > 0$ element-wise. Relatedly, if $v \leq w$ element-wise, and $v \neq w$, then $Pv < Pw$.

Let

$$Q = \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, x \neq 0\}$$

be the nonnegative orthant, excluding the origin. In addition, let

$$C = \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, \|x\| = 1\},$$

where $\|\cdot\|$ is any vector norm. Clearly, C is compact, i.e., closed and bounded.

Then, for all $z \in Q$, we can define the following function: let

$$f(z) = \max \{s \in \mathbb{R} : sz \leq Az\} = \min_{1 \leq i \leq n, z_i \neq 0} \frac{(Az)_i}{z_i}$$

Here are facts about f .

- $f(rz) = f(z)$, for all $r > 0$.
- If $Az = \lambda z$, i.e., if (λ, z) is an eigenpair, then $f(z) = \lambda$.
- If $sz \leq Az$, then $sPz \leq PAz = APz$, where the latter follows since A and P clearly commute. So,

$$f(z) \leq f(Pz).$$

In addition, if z is *not* an eigenvector of A , then $sz \neq Az$, for all s ; and $sPz < APz$. From the second expression for $f(z)$ above, we have that in this case that

$$f(z) < f(Pz),$$

i.e., an inequality in general but a strict inequality if not an eigenvector.

This *suggests* an idea for the proof: look for a positive vector that maximizes the function f ; show it is an eigenvector we want in the theorem; and show that it established the properties stated in the theorem.