

Lecture: Basic Matrix Results (3 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.

4 Review and overview

Recall the basic statement of the Perron-Frobenius theorem from last class.

Theorem 1 (Perron-Frobenius). *Let $A \in \mathbb{R}^{n \times n}$ be an irreducible non-negative matrix. Then,*

1. *A has a positive real eigenvalue $\lambda_{\max}$; which is equal to the spectral radius; and $\lambda_{\max}$ has an associated eigenvector x with all positive entries.*
2. *If $0 \leq B \leq A$, with $B \neq A$, then every eigenvalue σ of B satisfies $|\sigma| < \lambda_{\max} = \rho_A$. (Note that B does not need to be irreducible.) In particular, B can be obtained from A by zeroing out entries; and also all of the diagonal minors $A_{(i)}$ obtained from A by deleting the i^{th} row/column have eigenvalues with absolute value strictly less than $\lambda_{\max} = \rho_A$. Informally, this says: ρ_A increases when any entry of A increases.*
3. *That eigenvalue ρ_A has algebraic and geometric multiplicity equal to one.*
4. *If $y \geq 0$, $y \neq 0$ is a vector and μ is a number such that $Ay \leq \mu y$, then $y > 0$ and $\mu \geq \lambda_{\max}$; with $\mu = \lambda_{\max}$ iff y is a multiple of x . Informally, this says: there is no other non-negative eigenvector of A different than x .*
5. *If, in addition, A is primitive/aperiodic, then each other eigenvalue λ of A satisfies $|\lambda| < \rho_A$.*
6. *If, in addition, A is primitive/aperiodic, then*

$$\lim_{t \rightarrow \infty} \left(\frac{1}{\rho_A} A \right)^t = xy^T,$$

where x and y are positive eigenvectors of A and A^T with eigenvalue ρ_A , i.e., $Ax = \rho_A x$ and $A^T y = \rho_A y$ (i.e., $y^T A = \rho_A y^T$), normalized such that $x^T y = 1$.

Today, we will do three things: (1) we will prove this theorem; (2) we will also discuss periodicity/aperiodicity issues; (3) we will also briefly discuss the first connectivity/non-connectivity result for Adjacency and Laplacian matrices of graphs that will use the ideas we have developed in the last few classes.

Before proceeding, one note: an interpretation of a matrix B generated from A by zeroing out an entry or an entire row/column is that you can remove an edge from a graph or you can remove

a node and all of the associated edges from a graph. (The monotonicity provided by that part of this theorem will be important for making claims about how the spectral radius behaves when such changes are made to a graph.) This obviously holds true for Adjacency matrices, and a similar statement also holds true for Laplacian matrices.

4.1 Proof of the Perron-Frobenius theorem

We start with some general notation and definitions; then we prove each part of the theorem in turn.

Recall from last time that we let $P = (I + A)^n$ and thus P is positive. Thus, for every non-negative and non-null vector v , then we have that $Pv > 0$ element-wise; and (equivalently) if $v \leq w$ element-wise, and $v \neq w$, then we have that $Pv < Pw$. Recall also that we defined

$$\begin{aligned} Q &= \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, x \neq 0\} \\ C &= \{x \in \mathbb{R}^n \text{ s.t. } x \geq 0, \|x\| = 1\}, \end{aligned}$$

where $\|\cdot\|$ is any vector norm. Note in particular that this means that C is compact, i.e., closed and bounded. Recall also that, for all $z \in Q$, we defined the following function: let

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

Finally, recall several facts about the function f .

- $f(rz) = f(z)$, for all $r > 0$.
- If $Az = \lambda z$, i.e., if (λ, z) is an eigenpair, then $f(z) = \lambda$.
- In general, $f(z) \leq f(Pz)$; and if z is *not* an eigenvector of A , then $f(z) < f(Pz)$. (The reason for the former is that if $sz \leq Az$, then $sPz \leq PAz = APz$. The reason for the latter is that in this case $sz \neq Az$, for all s , and $sPz < APz$, and by considering the second expression for $f(z)$ above.)

We will prove the theorem in several steps.

4.2 Positive eigenvalue with positive eigenvector.

Here, we will show that there is a positive eigenvalue λ^* and that the associated eigenvector x^* is a positive vector.

To do so, consider $P(C)$, the image of C under the action of the operator P . This is a compact set, and all vectors in $P(C)$ are positive. By the second expression in definition of $f(\cdot)$ above, we have that f is continuous on $P(C)$. Thus, f achieves its maximum value on $P(C)$, i.e., there exists a vector $x \in P(C)$ such that

$$f(x) = \sup_{z \in C} f(Pz).$$

Since $f(z) \leq f(Pz)$, the vector x realizes the maximum value f_{max} of f on Q . So,

$$f_{max} = f(x) \leq f(Px) \leq f_{max}.$$

Thus, from the third property of f above, x is an eigenvector of A with eigenvalue f_{max} . Since $x \in P(C)$, then x is a positive vector; and since $Ax > 0$ and $Ax = f_{max}x$, it follows that $f_{max} > 0$.

(Note that this result shows that $f_{max} = \lambda^*$ is achieved on an eigenvector $x = x^*$, but it doesn't show yet that it is equal to the spectral radius.)

4.3 That eigenvalue equals the spectral radius.

Here, we will show that $f_{max} = \rho_A$, i.e., f_{max} equals the spectral radius.

To do so, let $z \in \mathbb{C}^n$ be an eigenvector of A with eigenvalue $\lambda \in \mathbb{C}$; and let $|z|$ be a vector, each entry of which equals $|z_i|$. Then, $|z| \in Q$.

We claim that $|\lambda||z| \leq A|z|$. To establish the claim, rewrite it as $|\lambda||z_i| \leq \sum_{k=1}^n A_{ik}|z_k|$. Then, since $Az = \lambda z$, i.e., $\lambda z_i = \sum_{k=1}^n A_{ik}z_k$, and since $A_{ik} \geq 0$, we have that

$$|\lambda||z| = \left| \sum_{k=1}^n A_{ik}z_k \right| \leq \sum_{k=1}^n A_{ik}|z_k|,$$

from which the claim follows.

Thus, by the definition of f (i.e., since $f(z) = \min \frac{(Az)_i}{(z)_i}$, we have that $|\lambda| \leq f(|z|)$. Hence, $|\lambda| \leq f_{max}$, and thus $\rho_A \leq f_{max}$ (where ρ_A is the spectral radius). Conversely, from the above, i.e., since f_{max} is an eigenvalue it must be $\leq$ the maximum eigenvalue, we have that $f_{max} \leq \rho_A$. Thus, $f_{max} = \rho_A$.

4.4 An extra claim to make.

We would like to establish the following result:

$$f(z) = f_{max} \Rightarrow (Az = f_{max}z \text{ and } z > 0).$$

To establish this result, observe that above it is shown that: if $f(z) = f_{max}$, then $f(z) = f(Pz)$. Thus, z is an eigenvector of A for eigenvalue f_{max} . It follows that $Pz = \lambda z$, i.e., that z is also an eigenvector of P . Since P is positive, we have that $Pz > 0$, and so z is positive.

4.5 Monotonicity of spectral radius.

Here, we would like to show that $0 \leq B \leq A$ and $B \neq A$ implies that $\rho_B < \rho_A$. (Recall that B need *not* be irreducible, but A is.)

To do so, suppose that $Bz = \lambda z$, with $z \in \mathbb{C}^n$ and with $\lambda \in \mathbb{C}$. Then,

$$|\lambda||z| \leq B|z| \leq A|z|,$$

from which it follows that

$$|\lambda| \leq f_A(|z|) \leq \rho_A,$$

and thus $\rho_B \leq \rho_A$.

Next, assume for contradiction that $|\lambda| = \rho_A$. Then from the above claim (in Section 4.4), we have that $f_A(z) = \rho_A$. Thus from above it follows that $|z|$ is an eigenvector of A for the eigenvalue ρ_A and also that z is positive. Hence, $B|z| = A|z|$, with $z > 0$; but this is impossible unless $A = B$.

Remark. Replacing the i^{th} row/column of A by zeros gives a non-negative matrix $A_{(i)}$ such that $0 \leq A_{(i)} \leq A$. Moreover, $A_{(i)} \neq A$, since the irreducibility of A precludes the possibility that all entries in a row are equal to zero. Thus, for all matrices $A_{(i)}$ that are obtained by eliminating the i^{th} row/column of A , the eigenvalues of $A_{(i)} < \rho$.

4.6 Algebraic/geometric multiplicities equal one.

Here, we will show that the algebraic and geometric multiplicity of λ_{max} equal 1. Recall that the geometric multiplicity is less than or equal to the algebraic multiplicity, and that both are at least equal to one, so it suffices to prove this for the algebraic multiplicity.

Before proceeding, also define the following: given a square matrix A :

- Let $A_{(i)}$ be the matrix obtained by eliminating the i^{th} row/column. In particular, this is a smaller matrix, with one dimension less along each column/row.
- Let A_i be the matrix obtained by zeroing out the i^{th} row/column. In particular, this is a matrix of the same size, with all the entries in one full row/column zeroed out.

To establish this result, here is a lemma that we will use; its proof (which we won't provide) boils down to expanding $\det(\Lambda - A)$ along the i^{th} row.

Lemma 1. *Let A be a square matrix, and let Λ be a diagonal matrix of the same size with $\lambda_1, \dots, \lambda_n$ (as variables) along the diagonal. Then,*

$$\frac{\partial}{\partial \lambda_i} \det(\Lambda - A) = \det(\Lambda_{(i)} - A_{(i)}),$$

where the subscript (i) means the matrix obtained by eliminating the i^{th} row/column from each matrix.

Next, set $\lambda_i = \lambda$ and apply the chain rule from calculus to get

$$\frac{d}{d\lambda} \det(\lambda I - A) = \sum_{i=1}^n \det(\lambda I - A_{(i)}).$$

Finally, note that

$$\det(\lambda I - A_i) = \lambda \det(\lambda I - A_{(i)}).$$

But by what we just proved (in the Remark at the end of last page), we have that $\det(\rho_A I - A_{(i)}) > 0$. Thus, the derivative of the characteristic polynomial of A is nonzero at ρ_A , and so the algebraic multiplicity equals 1.

4.7 No other non-negative eigenvectors, etc.

Here, we will prove the claim about other non-negative vectors, including that there are no other non-negative eigenvectors.

To start, we claim that: $0 \leq B \leq A \Rightarrow f_{\max}(B) \leq f_{\max}(A)$. (This is related to but a little different than the similar result we had above.) To establish the claim, note that if $z \in Q$ is s.t. $sz \leq Bz$, then $sz \leq Az$ (since $Bz \leq Az$), and so $f_B(z) \leq f_A(z)$, for all z .

We can apply that claim to A^T , from which it follows that A^T has a positive eigenvalue, call it η . So, there exists a row vector, $w > 0$ s.t. $w^T A = \eta w^T$. Recall that $x > 0$ is an eigenvector of A with maximum eigenvalue $\lambda_{\max}$. Thus,

$$w^T A x = \eta w^T x = \lambda_{\max} w^T x,$$

and thus $\eta = \lambda_{\max}$ (since $w^T x > 0$).

Next, suppose that $y \in Q$ and $Ay \leq \mu y$. Then,

$$\lambda_{\max} w^T y = w^T A y \leq \mu w^T y,$$

from which it follows that $\lambda_{\max} \leq \mu$. (This is since all components of w are positive and some components of y is positive, and so $w^T y > 0$).

In particular, if $Ay = \mu y$, then $\mu = \lambda_{\max}$.

Further, if $y \in Q$ and $Ay \leq \mu y$, then $\mu \geq 0$ and $y > 0$. (This is since $0 < Py = (I + A)^{n-1} y \leq (1 + \mu)^{n-1} y$.)

This proves the first two parts of the result; now, let's prove the last part of the result.

If $\mu = \lambda_{\max}$, then $w^T (Ay - \lambda_{\max} y) = 0$. But, $Ay - \lambda_{\max} y \leq 0$. So, given this, from $w^T (Ay - \lambda_{\max} y) = 0$, it follows that $Ay = \lambda_{\max} y$. Since y must be an eigenvector with eigenvalue $\lambda_{\max}$, the last result (i.e., that y is a scalar multiple of x) follows since $\lambda_{\max}$ has multiplicity 1.

To establish the converse direction march through these steps in the other direction.

4.8 Strict inequality for aperiodic matrices

Here, we would like to establish the result that the eigenvalue we have been talking about is strictly larger in magnitude than the other eigenvalues, under the aperiodicity assumption.

To do so, recall that the t^{th} powers of the eigenvalues of A are the eigenvalues of A^t . So, if we want to show that there does *not* exist eigenvalues of a primitive matrix with absolute value = ρ_A , other than ρ_A , then it suffices to prove this for a positive matrix A .

Let A be a positive matrix, and suppose that $Az = \lambda z$, with $z \in \mathbb{C}^n$, $\lambda \in \mathbb{C}$, and $|\lambda| = \rho_A$, in which case the goal is to show $\lambda < \rho_A$.

(We will do this by showing that any eigenvector with eigenvalue equal in magnitude to ρ_A is the top eigenvalue.) (I.e., we will show that such a z equals $|z|$ and thus there is no other one with ρ_A .)

Then,

$$\rho_A |z| = |Az| \leq A|z|,$$

from which it follows that

$$\rho_A \leq f(|z|) \leq \rho_A,$$

which implies that $f(|z|) = \rho_A$. From a result above, this implies that $|z|$ is an eigenvector of A with eigenvalue ρ_A . Moreover, $|Az| = A|z|$. In particular,

$$\left| \sum_{i=1}^n A_{1i} z_i \right| = \sum_{i=1}^n A_{1i} |z_i|.$$

Since all of the entries of A are positive, this implies that there exists a number $u \in \mathbb{C}$ (with $|u| = 1$) s.t. for all $i \in [n]$, we have that $z_i = u|z_i|$. Hence, z and $|z|$ are collinear eigenvectors of A . So, the corresponding eigenvalues of λ and ρ are equal, as required.

4.9 Limit for aperiodic matrices

Here, we would like to establish the limiting result.

To do so, note that A^T has the same spectrum (including multiplicities) as A ; and in particular the spectral radius of A^T equals ρ_A .

Moreover, since A^T is irreducible (a consequence of being primitive), we can apply the Perron-Frobenius theorem to it to get $yA = \rho_A y$. Here y is determined up to a scalar multiple, and so let's choose it s.t. $x^T y = \sum_{i=1}^n x_i y_i = 1$.

Next, observe that we can decompose the n -dimensional vector space $\mathbb{R}^n$ into two parts,

$$\mathbb{R}^n = R \oplus N,$$

where both R and N are invariant under the action of A . To do this, define the rank-one matrix $H = xy^T$, and:

- let R be the *image space* of H ; and
- let N be the *null space* of H .

Note that H is a projection matrix (in particular, $H^2 = H$), and thus $I - H$ is also a projection matrix, and the image space of $I - H$ is N . Also,

$$AH = Axy^T = \rho_A xy^T = x\rho_A y^T = xy^T A = HA.$$

So, we have a direct sum decomposition of the space $\mathbb{R}^n$ into $R \oplus N$, and this decomposition is invariant under the action of A .

Given this, observe that the restriction of A to N has all of its eigenvalues strictly less than ρ_A in absolute value, while the restriction of A to the one-dimensional space R is simply a multiplication/scaling by ρ_A . So, if P is defined to be $P = \frac{1}{\rho_A} A$, then the restriction of P to N has its eigenvalues < 1 in absolute value. This decomposition is also invariant under all positive integral powers of P . So, the restriction of P^k to N tends to zero as $k \rightarrow \infty$, while the restriction of P to R is the identity. So, $\lim_{t \rightarrow \infty} \left(\frac{1}{\rho_A} A \right)^t = H = xy^T$.

4.10 Additional discussion form periodicity/aperiodic and cyclicity/primitiveness

Let's switch gears and discuss the periodicity/aperiodic and cyclicity/primitiveness issues.

(This is an algebraic characterization, and it holds for general non-negative matrices. I think that most people find this less intuitive than the characterization in terms of connected components, but it's worth at least knowing about it.)

Start with the following definition.

Definition 1. *The cyclicity of an irreducible non-negative matrix A is the g.c.d. (greatest common denominator) of the length of the cycles in the associated graph.*

Let's let $\mathbb{N}_{ij}$ be a positive subset of the integers s.t.

$$\{t \in \mathbb{N} \text{ s.t. } (A^t)_{ij} > 0\},$$

that is, it is the values of $t \in \mathbb{N}$ s.t. the matrix A^t 's (i, j) entry is positive (i.e. exists a path from i to j of length t). Then, to define γ to be the cyclicity of A , first define $\gamma_i = \gcd(\mathbb{N}_{ii})$, and then clearly $\gamma = \gcd(\{\gamma_i \text{ s.t. } i \in V\})$. Note that each $\mathbb{N}_{ii}$ is closed under addition, and so it is a semi-group.

Here is a lemma from number theory (that we won't prove).

Lemma 2. *A set $\mathbb{N}$ of positive integers that is closed under addition contains all but a finite number of multiples of its g.c.d.*

From this it follows that $\forall i \in [n], \gamma_i = \gamma$.

The following theorem (which we state but won't prove) provides several related conditions for an irreducible matrix to be primitive.

Theorem 2. *Let A be an irreducible matrix. Then, the following are equivalent.*

1. *The matrix A is primitive.*
2. *All of the eigenvalues of A different from its spectral radius ρ_A satisfy $|\lambda| < \rho_A$.*
3. *The sequence of matrices $\left(\frac{1}{\rho_A} A\right)^t$ converges to a positive matrix.*
4. *There exists an $i \in [n]$ s.t., $\gamma_i = 1$.*
5. *The cyclicity of A equals 1.*

For completeness, note that sometimes one comes across the following definition.

Definition 2. *Let A be an irreducible non-negative square matrix. The period of A is the g.c.d. of all natural numbers m s.t. $(A^m)_{ii} > 0$ for some i . Equivalently, the g.c.d. of the lengths of closed directed paths of the directed graph G_A associated with A .*

Fact. All of the statements of the Perron-Frobenius theorem for positive matrices remain true for irreducible aperiodic matrices. In addition, all of those statements generalize to periodic matrices. The the main difference in this generalization is that for periodic matrices the “top” eigenvalue isn’t “top” any more, in the sense that there are other eigenvalues with equal absolute value that are different: they equal the p^{th} roots of unity, where p is the periodicity.

Here is an example of a generalization.

Theorem 3. *Let A be an irreducible non-negative $n \times n$ matrix, with period equal to h and spectral radius equal to $\rho_A = r$. Then,*

1. $r > 0$, and it is an eigenvalue of A .
2. r is a simple eigenvalue, and both its left and right eigenspace are one-dimensional.
3. A has left/right eigenvectors v/w with eigenvalue r , each of which has all positive entries.
4. A has exactly h complex eigenvalues with absolute value $= r$; and each is a simple root of the characteristic polynomial and equals the $r \cdot h^{\text{th}}$ root of unity.
5. If $h > 0$, then there exists a permutation matrix P s.t.

$$PAP^T = \begin{pmatrix} 0 & A_{12} & & 0 \\ & 0 & A_{23} & \\ & & \ddots & \ddots \\ 0 & & & 0 & A_{h-1,h} \\ A_{h1} & 0 & & & 0 \end{pmatrix}. \quad (1)$$

4.11 Additional discussion of directness, periodicity, etc.

Today, we have been describing Perron-Frobenius theory for non-negative matrices. There are a lot of connections with graphs, but the theory can be developed algebraically and linear-algebraically, i.e., without any mention of graphs. (We saw a hint of this with the g.c.d. definitions.) In particular, Theorem 3 is a statement about matrices, and it’s fair to ask what this might say about graphs we will encounter. So, before concluding, let’s look at it and in particular at Eqn. (1) and ask what that might say about graphs—and in particular undirected graphs—we will consider.

To do so, recall that the Adjacency Matrix of an undirected graph is symmetric; and, informally, there are several different ways (up to permutations, etc.) it can “look like.” In particular:

- It can look like this:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}, \quad (2)$$

where let’s assume that all-zeros blocks are represented as 0 and so each A_{ij} is not all-zeros. This corresponds to a vanilla graph you would probably write down if you were asked to write down a graph.

- It can look like this:

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \quad (3)$$

in which case the corresponding graph is not connected.

- It can even look like this:

$$A = \begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix}, \quad (4)$$

which has the interpretation of having two sets of nodes, each of which has edges to only the other set, and which will correspond to a bipartite graph.

- Of course, it could be a line-like graph, which would look like a tridiagonal banded matrix, which is harder for me to draw in latex, or it can look like all sorts of other things.
- But it *cannot* look like this:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad (5)$$

and it *cannot* look like this:

$$A = \begin{pmatrix} 0 & A_{12} \\ 0 & 0 \end{pmatrix}, \quad (6)$$

where recall we are assuming that each A_{ij} is not all-zeros. In both of these cases, these matrices are not symmetric.

In light of today's results and looking forward, it's worth commenting for a moment on the relationship between Eqns. (1) and Eqns (2) through (6).

Here are a few things to note.

- One might think from Eqns. (1) that periodicity means that that the graph is directed and so if we work with undirected graphs we can ignore it. That's true if the periodicity is 3 or more, but note that the matrix of Eqn (4) is periodic with period equal to 2. In particular, Eqn (4) is of the form of Eqn. (1) if the period $h = 2$. (It's eigenvalues are real, which they need to be since the matrix is symmetric, since the complex " 2^{th} roots of unity," which equal ± 1 , are both real.)
- You can think of Eqn. (3) as a special case of Eqn. (5), with the A_{12} block equal to 0, but it is not so helpful to do so, since its behavior is very different than for an irreducible matrix with $A_{12} \neq 0$.
- For directed graphs, e.g., the graph that would correspond to Eqn. (5) (or Eqn. (6)), there is very little spectral theory. It is of interest in practice since edges are often directed. But, most spectral graph methods for directed graphs basically come up—either explicitly or implicitly—with some sort of symmetrized version of the directed graph and then apply undirected spectral graph methods to that symmetrized graph. (Time permitting, we'll see an example of this at some point this semester.)
- You can think of Eqn. (5) as corresponding to a "bow tie" picture (that I drew on the board and that is a popular model for the directed web graph and other directed graphs). Although this is directed, it can be made irreducible by adding a rank-one update of the form $\mathbf{1}\mathbf{1}^T$ to the adjacency matrix. E.g., $A \rightarrow A + \epsilon\mathbf{1}\mathbf{1}^T$. This has a very natural interpretation in terms of random walkers, it is the basis for a lot of so-called "spectral ranking" methods, and it is a very popular way to deal with directed (and undirected) graphs. In addition, for reasons we will point out later, we can get spectral methods to work in a very natural way in this particular case, even if the initial graph is undirected.