Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg[*]

December 1, 2021 (unfinished!)

Contents

1. Positive and nonnegative matrices ([HorJoh, Chapter 8]) (cont'd) 1
1.1. The spectral radius (cont'd) 1
1.2. Perron–Frobenius theorems 5
1.2.1. Motivation 5
1.2.2. The theorems 6

Math 504 Lecture 24

1. Positive and nonnegative matrices ([HorJoh, Chapter 8]) (cont'd)

1.1. The spectral radius (cont'd)

Recall from last lecture:

The **spectral radius** of a square matrix A is

$$\rho(A) := \max\{|\lambda| \mid \lambda \in \sigma(A)\}.$$

Corollary 1.1.1. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ satisfy $B \geq A \geq 0$, then $\rho(A) \leq \rho(B)$.

(Inequalities between matrices are entrywise. Nonnegative or positive matrices have real entries by definition.)

Let us now prove some bounds for $\rho(A)$ when $A \geq 0$.

[*]Drexel University, Korman Center, 15 S 33rd Street, Philadelphia PA, 19104, USA

Definition 1.1.2. Let $\mathbb{F}$ be a field. Let $A \in \mathbb{F}^{n \times m}$.

(a) The **column sums** of A are the m sums

$$\sum_{i=1}^{n} A_{i,j} = (\text{the sum of all entries of the } j\text{-th column of } A) \qquad \text{for } j \in [m].$$

(b) The **row sums** of A are

$$\sum_{j=1}^{n} A_{i,j} = (\text{the sum of all entries of the } i\text{-th row of } A) \qquad \text{for } i \in [n].$$

(c) Now, assume that $\mathbb{F} = \mathbb{C}$ and $n > 0$ and $m > 0$. Then, we set

$$||A||_\infty := (\text{largest row sum of } |A|) = \max_{i \in [n]} \sum_{j=1}^{m} |A_{i,j}|$$

and

$$||A||_1 := (\text{largest column sum of } |A|) = \max_{j \in [m]} \sum_{i=1}^{n} |A_{i,j}|.$$

These two numbers are called the **∞-norm** and the **1-norm** of A (for reasons I will explain on zoom).

Example 1.1.3. The column sums of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are $a + c$ and $b + d$.

Warning 1.1.4. The column sums of a matrix are **not** the entries of the sum of its columns. Rather, the latter entries are the row sums, whereas the column sums are the entries of the sum of the rows.

Remark 1.1.5. Let $A \in \mathbb{F}^{n \times m}$.

(a) The row sums of A are the column sums of A^T.

(b) If $\mathbb{F} = \mathbb{C}$, then $||A||_\infty = \left|\left|A^T\right|\right|_1$.

Lemma 1.1.6. Let $A \in \mathbb{C}^{n \times n}$. Then:

(a) We have $\rho(A) \leq ||A||_\infty$.

(b) If $A \geq 0$ and if all row sums of A are equal, then $\rho(A) = ||A||_\infty$.

(c) We have $\rho(A) \leq ||A||_1$.

(d) If $A \geq 0$ and if all column sums of A are equal, then $\rho(A) = ||A||_1$.

Proof. **(a)** We have $\rho(A) = |\lambda|$ for some eigenvalue λ of A. Consider this λ, and let $v = (v_1, v_2, \ldots, v_n)^T \in \mathbb{C}^n$ be a nonzero λ-eigenvector.

Choose an $i \in [n]$ such that $|v_i| = \max\{|v_1|, |v_2|, \ldots, |v_n|\}$. Then, $|v_i| > 0$ (since v is nonzero).

Now, the i-th entry of the column vector Av is $\sum_{j=1}^{n} A_{i,j}v_j$; however, the same entry is λv_i (since $Av = \lambda v$). Comparing these facts, we obtain

$$\lambda v_i = \sum_{j=1}^{n} A_{i,j} v_j.$$

Taking absolute values, we obtain

$$\begin{aligned} |\lambda v_i| &= \left| \sum_{j=1}^{n} A_{i,j} v_j \right| \leq \sum_{j=1}^{n} \underbrace{\left| A_{i,j} v_j \right|}_{\substack{= \left|A_{i,j}\right| \cdot \left|v_j\right| \leq \left|A_{i,j}\right| \cdot |v_i| \\ \text{(since the choice of } i \\ \text{yields } \left|v_j\right| \leq |v_i| \text{)}}} \qquad \text{(by the triangle inequality)} \\ &\leq \sum_{j=1}^{n} \left| A_{i,j} \right| \cdot |v_i| . \end{aligned}$$

Since $|v_i| > 0$, we can cancel $|v_i|$ from this inequality (since the left hand side is $|\lambda v_i| = |\lambda| \cdot |v_i|$), and thus we obtain

$$\begin{aligned} |\lambda| &\leq \sum_{j=1}^{n} \left| A_{i,j} \right| = (\text{the } i\text{-th row sum of } |A|) \\ &\leq (\text{the largest row sum of } |A|) = ||A||_{\infty} . \end{aligned}$$

Since $\rho = |\lambda|$, this rewrites as $\rho \leq ||A||_{\infty}$, qed.

(b) Assume that $A \geq 0$ and that all row sums of A are equal. Let $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, and let κ be the common value of the row sums of A. Then, $Ae = \kappa e$ (since Ae is the vector whose entries are the row sums of A, but all these row sums are equal to κ). Hence, κ is an eigenvalue of A (since $e \neq 0$), so that $\rho(A) \geq |\kappa| = \kappa$ (since $A \geq 0$ entails $\kappa \geq 0$).

On the other hand, part **(a)** of the lemma yields

$$\rho(A) \leq ||A||_{\infty} = (\text{the largest row sum of } |A|) = (\text{the largest row sum of } A) = \kappa$$

(since all row sums of A are κ). Combining this with $\rho(A) \geq \kappa$, we obtain $\rho(A) = \kappa = ||A||_{\infty}$. This proves part **(b)**.

(c) Apply part **(a)** to A^T instead of A, and recall that $\left|\left|A^T\right|\right|_{\infty} = ||A||_1$ and $\rho\left(A^T\right) = \rho(A)$.

(d) Similar to part **(c)**. $\square$

Now, we can bound $\rho(A)$ from both sides when $A \geq 0$:

Theorem 1.1.7. Let $A \in \mathbb{R}^{n \times n}$ satisfy $A \geq 0$. Then,

$$(\text{the smallest row sum of } A) \leq \rho(A) \leq (\text{the largest row sum of } A).$$

Proof. The inequality $\rho(A) \leq (\text{the largest row sum of } A)$ follows from part **(a)** of the lemma, because

$$||A||_\infty = \left(\text{the largest row sum of } \underbrace{|A|}_{=A}\right) = (\text{the largest row sum of } A).$$

It remains to prove that $(\text{the smallest row sum of } A) \leq \rho(A)$.

Let $r_1, r_2, \ldots, r_n$ be the row sums of A. Let r_i be the smallest among them. We must then prove that $r_i \leq \rho(A)$. If $r_i = 0$, this is obvious. So WLOG assume that $r_i > 0$. Hence, all $r_1, r_2, \ldots, r_n$ are positive.

Let B the $n \times n$-matrix whose (u, v)-th entry is $\dfrac{r_i}{r_u} A_{u,v}$. So B is obtained from A by scaling each row such that the row sums all become r_i. Hence, the matrix B is ≥ 0 (since $A \geq 0$), and its row sums are all equal to r_i. Hence, part **(b)** of the above lemma (applied to B instead of A) yields

$$\begin{aligned} \rho(B) &= ||B||_\infty = (\text{the largest row sum of } |B|) \\ &= (\text{the largest row sum of } B) = r_i \end{aligned}$$

(since all row sums of B are r_i). However, for each $u \in [n]$, we have $\dfrac{r_i}{r_u} A_{u,v} \leq A_{u,v}$ (since $r_i \leq r_u$). In other words, $B \leq A$. Hence, the Corollary from last time (applied to B and A instead of A and B) yields $\rho(B) \leq \rho(A)$. Hence, $r_i = \rho(B) \leq \rho(A)$, qed. $\square$

Corollary 1.1.8. Let $A \in \mathbb{R}^{n \times n}$ satisfy $A \geq 0$. Let $x_1, x_2, \ldots, x_n$ be any n positive reals. Then,

$$\min_{i \in [n]} \sum_{j=1}^{n} \frac{x_i}{x_j} a_{i,j} \leq \rho(A) \leq \max_{i \in [n]} \sum_{j=1}^{n} \frac{x_i}{x_j} a_{i,j}.$$

Proof. Let $D = \operatorname{diag}(x_1, x_2, \ldots, x_n)$. Then, we can apply the previous theorem to DAD^{-1} instead of A, and notice that the row sums of DAD^{-1} are exactly the sums $\sum_{j=1}^{n} \frac{x_i}{x_j} a_{i,j}$ for $i \in [n]$. (And, of course, $\rho\left(DAD^{-1}\right) = \rho(A)$). $\square$

Remark 1.1.9. If the $x_1, x_2, \ldots, x_n$ in the above corollary are chosen appropriately, both of the inequalities can become equalities. (This follows from the Perron–Frobenius theorems further below.)

1.2. Perron–Frobenius theorems

We now come to the most important results about nonnegative matrices: the Perron–Frobenius theorems.

1.2.1. Motivation

Recall a standard situation in probability theorem. Consider a system (e.g., a slot machine) that can be in one of n possible **states** $s_1, s_2, \ldots, s_n$. Every minute, the system randomly changes states according to the following rule: If the system is in state s_i, then it changes to state s_j with probability $P_{i,j}$, where P is a (fixed, predetermined) nonnegative $n \times n$-matrix whose row sums all equal 1 (such a matrix is called **row-stochastic**). This is commonly known as a **Markov chain**.

Given such a Markov chain, one often wonders about its "steady state": If you wait long enough, how likely is the system to be in a given state?

Example 1.2.1. Let $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$. We encode the two states s_1 and s_2 as the basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$ of the vector space $\mathbb{R}^{1\times 2}$ (we work with row vectors here for convenience). Thus, a probability distribution on the set of states (i.e., a distribution of the form "state s_1 with probability a_1 and state s_2 with probability a_2") corresponds to a row vector $(a_1, a_2) \in \mathbb{R}^{1\times 2}$ satisfying $a_1 \geq 0$ and $a_2 \geq 0$ and $a_1 + a_2 = 1$.

If we start at state s_1 and let k minutes pass, then the probability distribution for the resulting state is $s_1 P^k$. More generally, if we start with a probability distribution $d \in \mathbb{R}^{1\times 2}$ and let k minutes pass, then the resulting state will be distributed according to dP^k. So we wonder: What is $\lim\limits_{k\to\infty} dP^k$ as $k \to \infty$? Does this limit even exist?

We can notice one thing right away: If $\lim\limits_{k\to\infty} dP^k$ exists, then this limit is a left 1-eigenvector of P, in the sense that it is a row vector y such that $yP = y$ (since $y = \lim\limits_{k\to\infty} dP^k = \lim\limits_{k\to\infty} dP^{k+1} = \left(\lim\limits_{k\to\infty} dP^k\right) P = yP$). Since it is furthermore a vector whose coordinates add up to 1 (because it is a limit of such vectors), this often allows us to explicitly compute it. In the above case, for example, we get

$$\lim_{k\to\infty} dP^k = \left(\frac{5}{6}, \frac{1}{6}\right).$$

But does this limit actually exist? Yes, in this specific example, but this isn't quite that obvious. Note that this limit (known as the **steady state** of the Markov chain) actually does not depend on the starting distribution d.

Does this generalize? Not always. Here are two bad examples:

- If $P = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then $\lim\limits_{k\to\infty} dP^k = d$ for each d, so the limits do depend on d.
- If $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, then $\lim\limits_{k\to\infty} dP^k$ does not exist unless $d = (0.5,\ 0.5)$, since in all other cases the sequence $\left(dP^k\right)_{k\geq 0}$ oscillates between (a_1, a_2) and (a_2, a_1).

Perhaps surprisingly, such cases are an exception. For **most** row-stochastic matrices P (that is, nonnegative matrices whose row sums all equal 1), there is a **unique** steady state (i.e., left 1-eigenvector), and it can be obtained as $\lim\limits_{k\to\infty} dP^k$ for any starting distribution d. To be more precise, this holds whenever P is positive (i.e., all $P_{i,j} > 0$). Some weaker assumptions also suffice.

More general versions of these facts hold even if we don't assume P to be row-stochastic, but merely require $P > 0$ (or $P \geq 0$ with some extra conditions). These will be the Perron and Perron–Frobenius theorems.

1.2.2. The theorems

Theorem 1.2.2 (Perron theorem)**.** Let $A \in \mathbb{R}^{n\times n}$ satisfy $A > 0$. Then:

(a) We have $\rho(A) > 0$.

(b) The number $\rho(A)$ is an eigenvalue of A and has algebraic multiplicity 1 (and therefore geometric multiplicity 1 as well).

(c) There is a unique $\rho(A)$-eigenvector $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{C}^n$ of A with $x_1 + x_2 + \cdots + x_n = 1$. This eigenvector x is furthermore positive. (It is called the **Perron vector** of A.)

(d) There is a unique vector $y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{C}^n$ such that $y^T A = \rho(A)\, y^T$ and $x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 1$. This vector y is also positive.

(e) We have

$$\left(\frac{1}{\rho(A)} A\right)^m \to xy^T \qquad \text{as } m \to \infty.$$

(f) The only eigenvalue of A that has absolute value $\rho(A)$ is $\rho(A)$ itself.

We will prove this next time.

Theorem 1.2.3 (Perron–Frobenius 1)**.** Let $A \in \mathbb{R}^{n\times n}$ satisfy $A \geq 0$. Then:

(a) The number $\rho(A)$ is an eigenvalue of A.

(b) The matrix A has a nonzero nonnegative $\rho(A)$-eigenvector.

To get stronger statements without requiring $A > 0$, we need two further properties of A.

Definition 1.2.4. Let $A \in \mathbb{R}^{n \times n}$ be an $n \times n$-matrix.

(a) We say that A is **reducible** if there exist two disjoint nonempty subsets I and J of $[n]$ such that $I \cup J = [n]$ and such that

$$A_{i,j} = 0 \qquad \text{for all } i \in I \text{ and } j \in J.$$

Equivalently, A is reducible if and only if there exists a permutation matrix P such that

$$P^{-1} A P = \begin{pmatrix} B & C \\ 0_{(n-r) \times r} & D \end{pmatrix} \qquad \text{for some } 0 < r < n \text{ and some } B, C, D.$$

(b) We say that A is **irreducible** if A is not reducible.

(c) We say that A is **primitive** if there exists some $m > 0$ such that $A^m > 0$.

Theorem 1.2.5 (Perron–Frobenius 2). Let $A \in \mathbb{R}^{n \times n}$ be nonnegative and irreducible. Then:

(a) We have $\rho(A) > 0$.

(b) The number $\rho(A)$ is an eigenvalue of A and has algebraic multiplicity 1 (and therefore geometric multiplicity 1 as well).

(c) There is a unique $\rho(A)$-eigenvector $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{C}^n$ of A with $x_1 + x_2 + \cdots + x_n = 1$. This eigenvector x is furthermore positive. (It is called the **Perron vector** of A.)

(d) There is a unique vector $y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{C}^n$ such that $y^T A = \rho(A) y^T$ and $x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 1$. This vector y is also positive.

(e) Assume furthermore that A is primitive. We have

$$\left(\frac{1}{\rho(A)} A \right)^m \to x y^T \qquad \text{as } m \to \infty.$$

(f) Assume again that A is primitive. The only eigenvalue of A that has absolute value $\rho(A)$ is $\rho(A)$ itself.

Remark 1.2.6. If A is the row-stochastic matrix P corresponding to a Markov chain, then:

- A is irreducible if and only if there is no set of states from which you cannot escape (except for the empty set and for the set of all states);
- A is primitive if and only if there is no "oscillation".