MATH 470 Independent Study in Matrix Theory: A Few Problems on Non-negative Matrices

Thomas Zhang

1 Inequalities and generalities

Key theorems and proofs

Theorem 1.1 *Let $A = [a_{ij}] \in M_n$ and $x = [x_i] \in \mathbb{C}^n$ be given*

1. $|Ax| \leq |A||x|$

2. *Suppose that A is non-negative and has a positive row. If $|Ax| = A|x|$, then there is some rotation $\theta \in [0, 2\pi]$ such that $e^{i\theta}x = |x|$*

3. *Suppose that x is positive. If $Ax = |A|x$, then $A = |A|$ (A is non-negative).*

Proof: The first assertion follows expanding $|Ax|$ and using triangle inequality. The second assertion also looks at $|Ax|$, except we're now given an equality in the triangle inequality:

$$|Ax|_i = \Big|\sum_{j=1}^n a_{ij}x_i\Big| = \sum_{j=1}^n |a_{ij}x_i| = \sum_{j=1}^n |a_{ij}|\,|x_i| = (|A|\,|x|)_i$$

Since each entry a_{ij} of A is non-negative, some facts about complex numbers tells us that actually

$$\Big|\sum_{j=1}^n a_{ij}x_i\Big| = e^{i\theta}\sum_{j=1}^n a_{ij}x_i = \sum_{j=1}^n |a_{ij}|\,|x_i|.$$

If we have a positive row in A, then expanding the sum of that row, we see that each $e^{i\theta}a_{ij}x_i = |a_{ij}|\,|x_i| = a_{ij}|x_i|$ for each i, i.e. $x = e^{i\theta}|x|$ as required. For the third assertion, since x is positive, we have $|A|x = \mathrm{Re}(|A|x) = \mathrm{Re}(Ax) = \mathrm{Re}(A)x$ such that $(A - \mathrm{Re}(A))x = 0$. However, $|A| - \mathrm{Re}(A) > 0$, and $x > 0$, so $|A| = \mathrm{Re}(A) \implies A = |A| \geq 0$.

Theorem 1.2 *Let $A, B \in M_n$ and suppose that B is non-negative. If $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$.*

Proof: We observe that $A^m \leq |A|^m \leq B^m$ for each $m = 1, 2, \ldots$. Observe that this implies $\|A^m\|_2 \leq \||A|^m\|_2 \leq \|B^m\|_2$, where $\|\cdot\|_2$ is the entrywise matrix 2-norm. An application of Gelfand's formula (letting $m \to \infty$) gets us the required inequality of spectral radius.

Theorem 1.3 *If $\tilde{A}$ is a principal submatrix of a non-negative matrix A, then $\rho(\tilde{A}) \leq \rho(A)$ $\implies$ $\max_i |a_{ii}| \leq \rho(A)$ $\implies$ $\rho(A) > 0$ if any main diagonal entry of A is positive.*

Proof: Observe that we can embed $\tilde{A}$ into a $n \times n$ matrix $\tilde{A}' = \left[\tilde{A} \bigoplus 0_{n-r}\right]$, where the additional entries are all 0. We can find an appropriate permutation similarity transformation on A such that $PAP^{-1} = \begin{bmatrix} \tilde{A} & * \\ * & * \end{bmatrix}$. Observe that $\rho(A) = \rho(PAP^{-1})$, and $(PAP^{-1} - \tilde{A}') \geq 0$ such that $\rho(\tilde{A}) = \rho(\tilde{A}') \leq \rho(PAP^{-1}) = \rho(A)$.

Theorem 1.4 *Let $A = [a_{ij}] \in M_n$ be non-negative. Then $\rho(A) \leq |||A|||_\infty := \max_i \sum_{j=1}^n a_{ij}$ and $\rho(A) \leq |||A|||_1 := \max_j \sum_{i=1}^n a_{ij}$. Note that if all row sums are equal, then $\rho(A) = |||A|||_\infty$, and if all column sums are equal, then $\rho(A) = |||A|||_1$. What does this say about stochastic matrices...*

Proof: The first set of claims are by the nature of the spectral radius ($\rho(A) \leq |||A|||$). If the row sums are equal, then $Ae = |||A|||_\infty e$, i.e. we have found an eigenvalue that achieves $|||A|||_\infty$. We can do something similar when column sums are equal: $e^\top A = |||A|||_1 e^\top$. This proves the second claim.

2 Positive matrices

Key theorems and proofs

Theorem 2.1 (Perron's big'un) *Let $A \in M_n$ be positive. Then*

1. $\rho(A) > 0$
2. *$\rho(A)$ is an eigenvalue of algebraic multiplicity* 1
3. *there is a unique real eigenvector $x > 0$ such that $Ax = \rho(A)x$ and $x_1 + \cdots + x_n = 1$*
4. *there is a unique real eigenvector $y > 0$ such that $y^\top A = \rho(A)y^\top$ and $\langle x, y\rangle = 1$*
5. *$(\rho(A)^{-1})^m \to xy^\top$ as $m \to \infty$.*

Proof: This will be a long one. The first claim is trivial because A is positive and hence cannot be 0.

For the third and fourth claims, we observe that if λ, x are an eigenpair of A such that $|\lambda| = \rho(A)$ (similarly, λ, y for left eigenvector), then we have from the positivity of A that $Ax = \lambda x$ such that $|Ax| = A|x| = |\lambda||x|$. This implies that $|\lambda| = \rho(A), |x|$ are also an eigenpair of A. Since we have found a positive eigenvector for eigenvalue $\rho(A)$, all that is left to do is re-scale it such that $\sum_i |x|_i = 1$. We can similarly re-scale $|y|$ such that $\langle x, y\rangle = 1$.

For the second claim, we have from above that if λ, x such that $|\lambda| = \rho(A)$, then $|x|$ is an eigenvector for $\rho(A)$. However, we also have from Theorem 1.1.2 that if $|Ax| = A|x|$, which is fulfilled because $|Ax| = |\lambda x| = \rho(A)|x| = A|x|$, then there is some $e^{i\theta}$ such that $e^{i\theta}x = |x|$. However, this implies that $|x| = cx$, such that $|x|$ belongs to the eigenspace associated with λ. Therefore, $\lambda = \rho(A)$. The same line of logic can be used to show that

$\rho(A)$ can only have geometric multiplicity 1, since all eigenvectors necessarily are of form $ce^{i\theta}|x|$. To show that it has algebraic multiplicity, we observe that the eigenvectors x, y we constructed above by definition fulfill $y^*x = 1 \neq 0$ and are both eigenvectors for $\rho(A)$. Therefore, $\rho(A)$ must have an algebraic multiplicity of 1.

The last claim follows from a nice lemma regarding eigenvectors and similarity transformations, so nice that I'm going to prove it here. We claim that $A = S \begin{bmatrix} \rho(A) & 0 \\ 0 & B \end{bmatrix} S^{-1}$, where $S = [x\ S_1]$ and $S^{-*} = [y\ Z_1]$. We have from above that x, y are right and left eigenvectors of A for the same eigenvalue and $y^*x = 1$. Now we consider $S_1 \in M_{n,n-1}$, whose columns are a basis for the orthogonal complement of y, and set $S = [x\ S_1]$. We now prove that S is invertible. Say $z = [z_1\ \beta^*]^*$ such that $Sz = 0$. We then get $0 = y^*Sz = z_1(y^*x) = z_1$ by construction of S_1. Therefore, $Sz = S_1\beta = 0$, but S_1 has full column rank by construction, so $\beta = 0$. S is invertible. Let $S^{-*} = [\tau\ Z_1]$. By the following,

$$I_n = S^{-1}S = \begin{bmatrix} \tau^* \\ Z_1^* \end{bmatrix} \begin{bmatrix} x & S_1 \end{bmatrix} = \begin{bmatrix} \tau^*x & \tau^*S_1 \\ Z_1^*x & Z_1^*S_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0_{n-1} \end{bmatrix},$$

we see that τ is orthogonal to every column in S_1, so it must belong to $span(y)$, and $\tau^*x = 1$ tells us that $\tau = y$. Updating $S^{-*} = [y\ Z_1]$, we see

$$\begin{bmatrix} y^*x & y^*S_1 \\ Z_1^*x & Z_1^*S_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0_{n-1} \end{bmatrix}.$$

Taking note of each block's value, we now look at the similarity transformation on A:

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} A \begin{bmatrix} x & S_1 \end{bmatrix} \\ &= \begin{bmatrix} y^*Ax & y^*AS_1 \\ Z_1^*Ax & Z_1^*AS_1 \end{bmatrix} \\ &= \begin{bmatrix} \rho(A)y^*x & \rho(A)y^*AS_1 \\ \rho(A)Z_1^*x & Z_1^*AS_1 \end{bmatrix} \\ &= \begin{bmatrix} \rho(A) & 0 \\ 0 & Z_1^*AS_1 \end{bmatrix} =: \begin{bmatrix} \rho(A) & 0 \\ 0 & B \end{bmatrix}. \end{aligned}$$

This completes the lemma. Now we have all we need to prove the last claim of the Perron-Frobenius theorem. Let us perform the above similarity transformation on A, and observe that since $\rho(A)$ has algebraic multiplicity 1, $\rho(B) < \rho(S^{-1}AS) = \rho(A)$, such that $\rho\left(\frac{1}{\rho(A)}B\right) \leq \rho\left(\frac{1}{\rho(A)}A\right) = 1$. Therefore, $(\rho(A)^{-1}B)^m \to 0, m \to \infty$.

$$\begin{aligned} (\rho(A)^{-1}A)^m = S \begin{bmatrix} 1 & 0 \\ 0 & (\rho(A)^{-1}B)^m \end{bmatrix} S^{-1} &= \begin{bmatrix} x & S_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (\rho(A)^{-1}B)^m \end{bmatrix} \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} \\ &\to \begin{bmatrix} x & S_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0_{n-1} \end{bmatrix} \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} \\ &= xy^* = xy^\top. \end{aligned}$$

Theorem 2.2 (Another Fan theorem) *Let $A = [a_{ij}] \in M_n$. Suppose that $B = [b_{ij}] \in M_n$ is non-negative and $b_{ij} \geq |a_{ij}|$ for non-diagonal entries. Then every eigenvalue of A is in the union of n discs*

$$\bigcup_{i=1}^{n} \{z \in \mathbb{C} : |z - a_{ii}| \leq \rho(B) - b_{ii}\}.$$

In particular, A is non-singular if $|a_{ii}| > \rho(B) - b_{ii}|$ for all $i = 1, \ldots, n$.

3 Non-negative matrices

Key theorems and proofs

General theme: most things from positive matrices are still true, except at the highest level of generality, one must now switch out "positive" with "non-negative" when it comes to things like eigenvectors and the like. The interesting theorems in this and the following section tell us when we can actually say certain things are positive for non-negative matrices.

Theorem 3.1 *Let A be non-negative. Suppose that there is a positive vector x and non-negative real number λ such that $Ax = \lambda x$ or $x^\top A = \lambda x^\top$. Then $\lambda = \rho(A)$.*

Theorem 3.2 *Suppose that $A \in M_n$ is non-negative and has a positive left eigenvector $A^\top y = \rho(A) y$ ($\lambda = \rho(A)$ from previous theorem), then*

1. *If $x \in \mathbb{R}^n$ is non-zero and $Ax \geq \rho(A)x$, then x is an eigenvector corresponding to $\rho(A)$.*

2. *If $A \neq 0$, then $\rho(A) > 0$ and every maximum modulus eigenvalue ($|\lambda| = \rho(A)$) is semi-simple (algebraic multiplicity = geometric multiplicity).*

4 Irreducible nonnegative matrices

Key theorems and proofs

General theme: Irreducible non-negative matrices are quite special in the sense that most results for positive matrices can be generalized to irreducible non-negative matrices with little effort. First, two useful lemmas:

Lemma 4.1 (one characterization of irreducible non-negativity) *If $A \in M_n$ is non-negative, then A is irreducible if and only if $(I + A)^{n-1} > 0$.*

Lemma 4.2 *$\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A \in M_n$. Then $\lambda_1 + 1, \ldots, \lambda_n + 1$ are the eigenvalues of $I + A$ and $\rho(I + A) \leq \rho(A) + 1$. If A is non-negative, then we can use previous results to easily establish $\rho(I + A) = \rho(A) + 1$. This lemma is helpful by converting statements regarding A into statements regarding the positive matrix $(I + A)^{n-1}$ thanks to the previous lemma.*

We find that all but the last assertions of the Perron-Frobenius theorem can be ported entirely intact to irreducible non-negative matrices, positivity and all:

Theorem 4.3 (Perron-Frobenius revamped) *Let $A \in M_n$ be positive. Then*

1. $\rho(A) > 0$
2. *$\rho(A)$ is an eigenvalue of algebraic multiplicity* 1
3. *there is a unique real eigenvector $x > 0$ such that $Ax = \rho(A)x$ and $x_1 + \cdots + x_n = 1$*
4. *there is a unique real eigenvector $y > 0$ such that $y^\top A = \rho(A)y^\top$ and $\langle x, y\rangle = 1$*

Theorem 4.4 *Let $A, B \in M_n$. Suppose that A is non-negative and irreducible, and $A \geq |B|$. Let $\lambda = e^{i\theta}\rho(B)$ be a maximum-modulus eigenvalue of B. If $\rho(A) = \rho(B)$, then there is some diagonal unitary matrix D such that $B = e^{i\theta}DAD^{-1}$.*

Proof: Let x be an eigenvector of B such that $Bx = \lambda x$. Then we have $|Bx| = |B|\,|x| = |\lambda||x| = \rho(B)|x| = \rho(A)|x|$. But we also know that $|B| \leq A$ such that $|B||x| = \rho(A)|x| \leq A|x|$. Since $|x| \geq 0$, this tell us that actually $A|x| = \rho(A)|x|$. Moreover, since A is irreducible, $|x| > 0$ by the revamped Perron-Frobenius theorem. Now we argue that since $(A - |B|)|x| = 0$, yet $(A - |B|) \geq 0$ and $|x| > 0$, in fact $A = |B|$.

We find diagonal unitary matrix D such that $x = D|x|$. We then get $Bx = \lambda x = e^{i\theta}\rho(A)x$ for some $\theta \in [0, 2\pi)$. Therefore, $Bx = BD|x| = e^{i\theta}\rho(A)D|x| \implies e^{-i\theta}D^{-1}BD|x| = \rho(A)|x| = A|x| = |B||x|$. Setting $C = e^{-i\theta}D^{-1}BD$, since $C|x| = |C||x|$, we know that $C = |C| = |B| = A$. Therefore, $B = e^{i\theta}DAD^{-1}$ as required.

Corollary 4.5 *Let $A \in M_n$ be irreducible and non-negative, and suppose that it has precisely k distinct eigenvalues of maximum modulus. Then,*

1. *A is similar to $e^{2\pi ip/k}A$ for each $p = 0, 1, \ldots, k-1$*
2. *the maximum modulus eigenvalues of A are $e^{2\pi ip/k}\rho(A), p = 0, 1, \ldots, k-1$, each of algebraic multiplicity 1.*

Corollary 4.6 *Suppose that $A \in M_n$ is irreducible and non-negative. If A has $k > 1$ maximum modulus eigenvalues, then every main diagonal entry of A is zero. Moreover, every main diagonal entry of A^m is 0 for each positive integer m that is not divisible by k.*

Proof: From the previous corollary, we know that the k maximum modulus eigenvalues of A are of form $e^{2\pi ip/k}\rho(A)$. Letting $\theta = 2\pi/k$, we see that A is similar to $e^{i\theta}A$. This implies that $\operatorname{tr}(A) = e^{i\theta}\operatorname{tr}(A)$. Since $e^{i\theta}$ is emphatically not real, let alone positive, this equality can only hold if all the main diagonal elements of A are 0 such that $\operatorname{tr}(A) = 0$. Similarly, we find that A^n is similar to $e^{in\theta}A$ for any $n = 1, 2, \ldots$, and therefore $\operatorname{tr}(A^n) = e^{in\theta}\operatorname{tr}(A^n)$. If k does not divide n, then $e^{in\theta}$ is not real, so again the equality can only hold if every diagonal element of A^n is 0.

5 Primitive matrices

Key theorems and proofs

One can see that many theorems in the previous section mention the existence of multiple maximum modulus eigenvalues. What happens when we only look at irreducible non-negative matrices with a unique maximum modulus eigenvalue. We call these primitive matrices.

Theorem 5.1 *In addition to non-negativity, if we assume primitiveness, we can assert the last part of the Perron-Frobenius theorem that we are not able to in general for non-negative matrices: if $A \in M_n$ is primitive, and if x and y are the right and left Perron vectors of A, then $\lim_{n\to\infty}(\rho(A)^{-1}A)^m = xy^\top > 0$.*

Theorem 5.2 *If $A \in M_n$ is non-negative, then A is primitive if and only if $A^m > 0$ for some $m \geq 1$.*

Corollary 5.3 *If $A \in M_n$ is non-negative and primitive, then A^m is non-negative and primitive for all $m \geq 1$ because if we find $p \in \mathbb{N}$ such that $A^p > 0$, then certainly $A^{mp} > 0$.*

Theorem 5.4 (another way to identify primitivity) *If $A \in M_n$ is irreducible and non-negative, and if all its main diagonal entries are positive, then $A^{n-1} > 0$, so A must be primitive.*

The following are some bounds on finding m such that $A^m > 0$ for primitive A.

Theorem 5.5 (Wielandt) *Let $A \in M_n$ be non-negative. Then A is primitive if and only if $A^{n^2-2n+2} > 0$.*

Theorem 5.6 *Let $A \in M_n$ be irreducible and non-negative, and suppose A has d positive main diagonal entries. Then $A^{2n-d-1} > 0$.*

6 Stochastic and doubly stochastic matrices

Key theorems and proofs

Stochastic matrices are great. By their definition ($Ae = e$ or $e^\top A = e^\top$), we see that we immediately have a positive eigenvector, which also tells us $\rho(A) = 1$. Therefore, the family of stochastic matrices all share their largest eigenvalue as well as the corresponding eigenvector. For now, our interest in doubly stochastic matrices is directed to the Birkhoff decomposition of doubly stochastic matrices.

Theorem 6.1 *A matrix $A \in M_n$ is doubly stochastic if and only if there are permutation matrices $P_1, \ldots, P_n \in M_n$ and $\lambda_1, \ldots, \lambda_N \geq 0, \sum_{i=1}^N \lambda_i = 1$ such that*

$$A = \lambda_1 P_1 + \cdots + t_N P_N.$$

Furthermore, $N \leq n^2 - n + 1$. (We can actually reduce this to $n^2 - 2n + 2$ using a linear programming perspective and Caratheodory's theorem).

Corollary 6.2 *The maximum/minimum of a convex/concave real-valued function on the set of doubly stochastic matrices is attained at a permutation matrix.*

Lemma 6.3 *$A \in M_n$ is a doubly sub-stochastic matrix. Then there is some doubly stochastic $S \in M_n$ such that $A \leq S$.*

Now we have the sufficient tools to prove the square case of von Neumann's Trace Inequality:

Theorem 6.4 *Let the ordered singular values of $A, B \in M_n$ be $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$ and $\sigma_1(B) \geq \cdots \geq \sigma_n(B)$. Then*

$$\mathit{Re}\ \mathit{tr}(AB) \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B)$$

Proof: The first part of the proof is a simple manipulation using the properties of trace. Let $A = U_1 \Sigma_A V_1^*$ and $B = U_2 \Sigma_B V_2^*$ be the singular value decompositions of A and B such that $\Sigma_A = \operatorname{diag}(\sigma_1(A), \ldots, \sigma_n(A))$, $\Sigma_B = \operatorname{diag}(\sigma_1(B), \ldots, \sigma_n(B))$. Then, setting $U = V_1^* U_2$, $V = V_2^* U_1$, we can move things around in the trace to get:

$$\begin{aligned}
\operatorname{tr}(AB) &= \operatorname{tr}(U_1 \Sigma_A V_1^* U_2 \Sigma_B V_2^*) \\
&= \operatorname{tr}(\Sigma_A V_1^* U_2 \Sigma_B V_2^* U_1) \\
&= \operatorname{tr}(\Sigma_A U \Sigma_B V) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) u_{ij} v_{ji}.
\end{aligned}$$

We observe that $U = [u_{ij}]$ and $V = [v_{ij}]$ are unitary matrices, so if we consider the matrix $W = [|u_{ij} v_{ji}|]$, we observe that applying the AM-GM inequality gets us: $W = [|u_{ij} v_{ji}] = [(|u_{ij}|^2 |v_{ji}|^2)^{1/2}] \leq [(|u_{ij}|^2 + |v_{ji}|^2)/2]$. Since U and V are unitary, the last matrix in the inequality is doubly stochastic. Therefore, W is doubly substochastic. Now we go back to our trace inequality and apply our newfound knowledge:

$$\begin{aligned}
\operatorname{Re}\operatorname{tr}(AB) &\leq \operatorname{Re} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) u_{ij} v_{ji} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) \operatorname{Re}(u_{ij} v_{ji}) \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) |u_{ij} v_{ji}| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) s_{ij} \text{ where } S \text{ is doubly stoch.}
\end{aligned}$$

We can certainly define a function on the doubly stochastic matrices $f(S) := \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) s_{ij}$, where

$$\operatorname{Re}\operatorname{tr}(AB) \leq \sup_S f(S) \leq \sup_S \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B) s_{ij}.$$

We can now invoke Birkhoff's Theorem. $f(S)$ is a linear function with respect to S, and is therefore convex. By Birkhoff's Theorem, we f attains its maximum at a permutation matrix, such that

$$\mathrm{Re}\ \mathrm{tr}(AB) \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B)p_{ij},$$

where $p_{ij} = 1$ if and only if $j = \pi(i)$ and is 0 everywhere else. Therefore, we can collapse one of the summation symbols to get

$$\begin{aligned} \mathrm{Re}\ \mathrm{tr}(AB) &\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_i(A)\sigma_j(B)p_{ij} \\ &\leq \sum_{i=1}^{n} \sigma_i(A)\sigma_{\pi(i)}(B). \end{aligned}$$

Using a basic fact of majorization, we complete the desired inequality:

$$\mathrm{Re}\ \mathrm{tr}(AB) \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_{\pi(i)}(B) \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B).$$