Non-negative matrices

A. Eremenko

March 18, 2023

In this section we use the following terminology and notation: a square matrix A is called *non-negative* (*positive*) if all entries are non-negative (positive). This will be denoted by $A \geq 0$ or $A > 0$. Same notation will be applied to vectors.

Warning. In other sections of this course, notation $A \geq 0$, $A > 0$ may have completely different meaning!

For a square matrix A we denote by $\rho(A)$ the *largest absolute value of an eigenvalue:*

$$\rho(A) := \max\{|\lambda| : Ax = \lambda x \text{ for some } x \neq 0\}.$$

Perron's Theorem. *Suppose that $A > 0$. Then:*
a) $p = \rho(A)$ is an eigenvalue.
b) The eigenspace corresponding to p is one-dimensional, and contains an eigenvector $x > 0$.
c) For all other eigenvalues $\lambda \neq p$ we have $|\lambda| < p$.
d) If λ is an eigenvalue different from p, then there is no non-negative eigenvector corresponding to λ.
e) p is a simple root of the characteristic equation.

Proof. Consider the set S consisting of all $t \geq 0$ such that $Ax \geq tx$ for some $x \geq 0$.

Let us check that the set S contains a maximal element.

This set S is not empty, moreover it contains some positive number. Indeed take $x = (1, 1 \ldots, 1)^T$. Then $Ax > 0$ because $A > 0$, so one can find $t > 0$ such that $Ax \geq tx$.

The set S is bounded. Indeed, it is easy to see that for $t \in S$ we have $t \leq n \max\{a_{i,j}\}$, where n is the size of A.

The set is closed: if $Ax_n = t_n x_n$ and $t_n \to t_\infty$, we can always assume that $\|x_n\| = 1$, then take a subsequence such that $x_n \to x_\infty \geq 0$, $x \neq 0$. We will have $Ax_\infty \geq t_\infty x_\infty$ and $x_\infty \neq 0$, because $\|x_\infty\| = 1$, so the limit t_∞ is in S.

Thus by a theorem from Analysis, there is a maximum element of S, and we denote it by p.

Now we show that p is an eigenvalue. We have $Ax \geq px$ for some $x \geq 0$, $x \neq 0$. So $y := Ax - px$ is a non-negative vector. Suppose that $y \neq 0$. Then $Ay > 0$, because A is positive, and $Ax > 0$, so

$$AAx > pAx,$$

and we can slightly increase p, which contradicts its definition. So $y = 0$ and we proved that p is an eigenvalue.

Now consider any eigenvalue λ of A. Then $Ax = \lambda x$ for some $x \neq 0$. Denote by $|x|$ the vector with cpomponents $|x_j|$. Then

$$A|x| \geq |Ax| = |\lambda x| = |\lambda||x|. \tag{1}$$

So $\lambda \leq p$ by the definition of p. This proves that p is in fact the eigenvalue of maximal modulus so we have a).

As we have $Ax = px$ for some $x \geq 0$ and $x \neq 0$, we conclude that $x > 0$ because $Ax > 0$ and $p > 0$. To prove b), it remains to show that there cannot be another linearly independent eigenvector corresponding to p. Let y be one. We may assume that y is real (because $A - pI$ is a real matric so its null space has a real basis). Then $x + cy$ is also an eigenvector corresponding to eigenvalue p, which is linearly independent of x. Then we can choose a real c, so that $x + cy$ is non-negative, and non-zero, but one of its coordinate is 0. Indeed, all coordinates of x are positive, so for small $|c|$ the coordinates of $x + cy$ are also positive. They cannot remain positive for all real c. So there is a limiting value of c (positive or negative) for which one coordinate becomes zero for the first time. For this c the vector $x + cy$ is not zero because x and y are linearly independent, so it is non-negative, non-zero, with one zero coordinate. But as $A(x + cy) = p(x + cy)$, we obtain a contradiction because all coordinates of $A(x + cy)$ must be positive. This completes the proof of b).

Now we want to prove c): that for each eigenvalue $\lambda \neq p$ we actually have $|\lambda| < p$. (We already know from a) that $|\lambda| \leq p$.) Proving this by contradiction, suppose that $|\lambda| = p$ but $\lambda \neq p$. We proved above that $Ay \geq py$, $y \geq 0$ implies $Ay = py$. Applying this to (1) with $y = |x|$ we conclude that (1) must hold with equality: $A|x| = |\lambda||x|$. In particular, $|Ax| = A|x|$, that is

$$\sum_j a_{i,j}|x_j| = \left|\sum_j a_{i,j}x_j\right|,$$

which means that all x_j must have the same argument θ. Multiplying them on $e^{-i\theta}$ we obtain a new vector $z \geq 0$ such that $Az = \lambda z = |\lambda||z|$. So $\lambda = |\lambda| = p$, contrary to our assumption. The contradiction shows that $|\lambda| < p$, which is c).

To prove d), notice that A and A^T have the same characteristic polynomial. We proved that A has a positive eigenvalue p which has maximim modulus among all eigenvalues, and that there is a positive eigenvector corresponding to it. Then the same applies to A^T. This means that there is a *left eigenvector*, a row wector w satisfying

$$wA = pw, \quad w > 0.$$

Now if u is any (right) eigenvector corresponding to some eigenvalue λ, then we have we have

$$p(w,u) = pwu = wAu = \lambda(w,u).$$

As $\lambda \neq p$ we conclude that $wu = 0$, and since $u \neq 0$, vector u cannot be non-negative because w is positive. This proves d).

The proof of e) requires Jordan's theorem. Suppose by contradiction that p is a multiple root of the characteristic equation. As we proved that the eigenspace of p has dimension 1 (see b)), we conclude that there must be a generalized eigenvector y. So we have $x > 0$, $Ax = px$ and

$$Ay - py = xs \tag{2}$$

with some y. We can always choose this y to be real. Moreover, we can add to y any multiple of x, and y will still be a generalized eigenvector. Adding cx with large possitive c, we can make y positive. Equation (2) with $x > 0$ implies that $Ay > py$, and y is positive, this contradicts the definition of p. This proves e) and completes the proof of the theorem.

The eigenvalue $\rho(A)$ is called the Perron eigenvalue, and a corresponding positive eigenvector is called Perron's eigenvector.

A square matrix is called *stochastic* if $A \geq 0$ and all column sums are equal to 1.

Theorem. *For a positive stochastic matrix* $\rho(A) = 1$*. If* x *is any non-zero, non-negative vector, then*

$$\lim_{m \to \infty} A^m x = cv,$$

where v *is a Perron eigenvector and* $c > 0$*.*

Proof. Let $w = (1, 1, \ldots, 1)$. The condition that A is stochastic can be written as

$$wA = w.$$

Applying d) of Perron's theorem to A^T we conclude that $\rho(A^T) = \rho(A) = 1$ (see the proof of d)). To prove the second statement we expand x into linear combination of eigenvectors and generalized eigenvectors:

$$x = c_1 v_1 + \ldots + c_n v_n.$$

Assume tat v_1 is a Perron eigenvector. Now apply a high power of A. Then we have $A^m v_1 = v_1$. For genuine eigenvector v_j we have $A^m v_j = \lambda_j^m v_j \to 0$, because $|\lambda_j| < 1$. So if all v_j are genuine, we have

$$A^m x \to c_1 v_1. \tag{3}$$

To see that $c_1 > 0$, multiply both sides of (3) on $w = (1, 1, ..., 1)$ from the left and use $wA = w$.

Finally, if some of the v_j are generalized, a similar argument can be performed to prove that $A^m v_j \to 0$. For example, if $Av = \lambda v$ and $Au = \lambda u + v$, then $A^2 u = \lambda^2 u + 2\lambda v$, $A^3 u = \lambda^3 u + 3\lambda^2 v$, and so on.