Markov chains and the Frobenius-Perron theorem

1. The Perron-Frobenius theory of nonnegative matrices (based on notes by Mike Boyle)

A matrix of real numbers is called nonnegative if all the entries are nonnegative. It is called positive if all the entries are positive.

A matrix is called primitive if it is square, nonnegative, and some power of the matrix is positive.

Theorem (Perron theorem) If A is a primitive matrix, then A has a unique eigenvalue \( \lambda \) of greatest absolute value. \( \lambda \) is real, positive, of algebraic multiplicity 1, and has an associated eigenvector all of whose coordinates are positive. (This condition has a simple interpretation in terms of the graph \( G=(V,E) \) where \( V=\{v_1,\ldots,v_n\} \) and \( E=\{e_{ij} : v_i \rightarrow v_j \} \).

Lemma Let \( \Phi : V \rightarrow \mathbb{R}^d \) be a linear transformation of a finite-dimensional \( \mathbb{R}^d \) vector space. Let \( E \subseteq V \) be a polytope (= bounded set which is an intersection of a finite number of closed half-spaces and has a nonempty interior). Assume that \( T^k(E) \) is contained in the interior of \( E \). Then all eigenvalues of \( T \) have absolute value less than \( 1 \).

Proof: Clearly, it's enough to prove this in the case \( k=1 \), i.e., we assume \( T(E) \subseteq E \). Because \( T(E) \subseteq E \), any eigenvalue \( \lambda \) satisfies \( |\lambda| \leq 1 \), so we need to explain why it can't happen that \( |\lambda| = 1 \). First, if \( \lambda = e^{2\pi i/n} \) is a root of unity, then 1 is an eigenvalue of \( T^k \), so there is a vector \( v \) on the boundary of \( E \) s.t. \( T^k(v) \in E \), in contradiction to the assumption that \( T(E) \subseteq E \).
Second, if \( \lambda < e^\theta \) but \( \lambda \) is not a root of unity, we know from linear algebra that there are vectors \( u, v \) such that
\[
T(u) = \cos \theta u + \sin \theta v \\
T(v) = -\sin \theta u + \cos \theta v,
\]
(i.e., \( \theta \) is irrational).

T acts as a rotation by \( \theta \) on \( \text{span}(u, v) \). In this case, if we take \( w \in \text{span}(u, v) \) then the powers of \( T \) acting on \( w \), \( (T^k(w))_{k \geq 0} \), will have \( w \) as an accumulation point. This is impossible since these points are in the compact set \( T(E) \subseteq E \).

\[ \square \]

**Proof of the Perron theorem.** First, we show there is an eigenvector with positive coordinates. Let \( \Delta \) denote the standard simplex \( \{ v = (x_1, \ldots, x_n) : x_i \geq 0, \sum x_i = 1 \} \).

Define a function \( F : \Delta \to \Delta \) by \( F(v) = \frac{A \cdot v}{\sum (A \cdot v)} \). (Note that \( A \cdot v \) is nonnegative and has at least one positive coordinate - otherwise \( A \) has a row of zeros and \( A^k \) cannot be positive for any \( k \).) Since \( F \) is continuous, there exists \( v \in \Delta \) s.t. \( F(v) = v \) (by the Brouwer fixed point theorem - is there a more constructive argument?),

so \( v \) is an eigenvector: \( Av = \lambda v \). By the assumption that \( A \) is primitive,

\[ v = \frac{1}{\lambda} \lambda^k v \] so all of the coordinates of \( v \) are positive.

Next, we reduce the verification of the remaining claims to the case when \( \lambda = 1 \) and \( v = (1, 1, \ldots, 1) \). Note that the eigenvalue equation \( \sum_{j=1}^{n} a_{ij} v_j = \lambda v_i \) can be rewritten as \( \sum_{j=1}^{n} \frac{a_{ij} v_j}{\lambda v_i} = 1 \). So, if we define a new matrix

\[ B = (b_{ij})_{i,j \in \mathbb{N}} \]

by \( b_{ij} = \frac{a_{ij} v_j}{\lambda v_i} \), then \( v = (1, 1, \ldots, 1) \) is an eigenvector of \( B \) with eigenvalue 1. Furthermore, it is easy to see that \( B \) is still a primitive matrix (think of the graph path interpretation).

Finally note that any number \( \mu \) is an eigenvalue of \( A \) if and only if \( \frac{1}{\lambda} \) is an eigenvalue of \( B \) : If \( w = (w_1, w_2, \ldots) \) satisfies \( \sum_{j=1}^{n} a_{ij} w_j = \mu \cdot w_i \), then we can write

\[ \lambda \frac{w_i}{v_i} = \sum_{j=1}^{n} \frac{a_{ij} w_j}{\lambda v_i} = \sum_{j=1}^{n} b_{ij} \cdot \frac{w_j}{v_j} \]

It remains to show that in the case \( \lambda = 1 \), \( v = (1, 1, \ldots, 1) \), all other eigenvalues
satisfy $|x| < 1$. The trick is to now consider the action of $A$ on row vectors, not column vectors: since $A(e_i) = (e_i)$, it follows that for any $v = (v_1, ..., v_n) \in \Delta$, $vA \in \Delta$, since $\langle vA, (e_i) \rangle = v_ia_i \geq 0$. So $A$ acting from the left maps $\Delta$ to itself, and by the primitivity assumption, some power $A^k$ maps $\Delta$ to its interior.

By the earlier part of the proof, there is a left-eigenvector $v = (v_1, ..., v_n) \in \Delta$ with positive coordinates such that $VA = v$. Thus if we define $V = \{w = (w_1, ..., w_n) \in \mathbb{R}^n : \sum w_j = 1\}$, $E = \Delta - v$, then $V$ is an $A$-invariant subspace of dimension $n-1$, and $E \in V$ is a polytope such that $A^k(E) \subset E$ for some $k$. Applying Lemma 1, we see that all eigenvalues of $A|_V$ satisfy $|\lambda| < 1$, which is just what we needed.

Corollary 1: If $A$ is a primitive matrix and $v$ is a nonnegative eigenvector of $A$ with eigenvalue $\lambda$, then $\lambda$ is the eigenvalue of largest absolute value (so $v$ is a scalar multiple of the positive eigenvector from the theorem).

Proof: By primitivity, $A^k v$ is positive for some $k$, so $v$ is positive and $\lambda \geq 0$. If $v$ is the eigenvector corresponding to the eigenvalue $\lambda$ of largest absolute value, and we choose $t$ such that $v < tu$, then for any $n \geq 1$, $A^n v = A^n u < A^n u = \lambda^n u$. This cannot happen if $\lambda < \lambda$, so $\lambda = \lambda$. □
A square matrix $A$ is called **irreducible** if for any $i,j$ there is a $k \geq 1$ s.t. $(A^k)_{ij} > 0$. This means the graph associated with the matrix is (strongly) connected. It is a weaker condition than primitivity. The **period** of an irreducible matrix is the gcd of all $k$ such that $(A^k)_{ij} > 0$.

Suppose that $A$ is irreducible with period $p$. Fix some vertex $v_0$. For $0 \leq i \leq p-1$ define $C_i = \{ \text{vertex } u : \text{there is a path of length } i \pmod{p} \text{ from } v_0 \text{ to } u \}$. The sets $C_0, C_1, \ldots, C_{p-1}$ partition the vertex set $V$. All outgoing edges from a vertex in $C_i$ lead to a vertex in $C_{i+1 \pmod{p}}$. By relabelling the row and column indices of $A$, we can bring it to the form of a block matrix

$$
A = \begin{pmatrix}
0 & A_1 & 0 & \cdots & 0 \\
0 & 0 & A_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & A_{p-1} \\
A_p & 0 & \cdots & 0 & 0 \\
\end{pmatrix}
$$

**Proposition:** Let $A$ be a square nonnegative matrix. Then $A$ is primitive if and only if it is irreducible with period one.

**Proof:** If $A$ is primitive then it is irreducible, and the period must be 1 (otherwise $A^k$ will have blocks of $0$'s). Conversely, if $A$ is irreducible with period 1, using the exercise below it is not difficult to see that for large enough $k$ there will be a path in the graph associated with $A$ between any two vertices. \qed
Prove that if

Exercise: $n_1, n_2, \ldots, n_m$ are positive integers with $\gcd(n_1, \ldots, n_m) = 1$, then the set of linear combinations of the form $\sum_{j=1}^{m} a_j n_j$, where $a_1, \ldots, a_m$ are nonnegative integers, contains all sufficiently large integers.

Theorem 2 (The Perron-Frobenius Theorem): Let $A$ be an irreducible matrix with period $p$. Then $A$ has a nonnegative right eigenvector $v$, unique up to multiplication by a scalar. The corresponding eigenvalue $\lambda$ is the spectral radius. $v$ has only positive coordinates, and has the following structure: if we write $v = \left(\begin{array}{c} v^0 \\ v^1 \\ \vdots \\ v^{p-1} \end{array}\right)$ where $v^i$ is the vector corresponding to the coordinates in the component $C(i)$ defined above, then $v^0$ is the positive eigenvector of the matrix $A_{p,1} A_{1,2} A_{2,3} \cdots$ and $v^j = A_j v^{j-1}$.

(Note: it is easy to see that $A_{p,1} A_{1,2} A_{2,3} \cdots$ is a primitive matrix.)

Proof: The vector described in the theorem is clearly a positive eigenvector. Conversely, if $u = \left(\begin{array}{c} u^0 \\ u^1 \\ \vdots \\ u^{p-1} \end{array}\right)$ is a nonnegative eigenvector, then $A u^0 = u^1$, $A u^1 = u^2, \ldots, A_{p-1} u^{p-2} = u^0, A_p u^{p-1} = \lambda u^0$, so $A_{p-1} u_{p-2} u^0 = \lambda u^0$. This shows that $u$ is determined up to scalar multiplication. The other claims are left as an exercise.

Corollary 2: Powers of a primitive matrix: Suppose $A = (a_{ij})_{ij \in \mathbb{N}}$ is a primitive matrix, and $u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n)$ are (respectively) a positive left eigenvector and a positive right eigenvector for the spectral radius $\lambda$, normalized so that $\langle u, v \rangle = \lambda u v = 1$. Then
\( \frac{1}{\lambda} A^k \) converges exponentially fast to the matrix \( u v^T \). That is, for any \( i \in [n] \) we have
\[
\frac{1}{\lambda} (A^k)_{ij} \xrightarrow[k \to \infty]{} u_i v_j
\]
Alternatively, we can write
\[
(A^k)_{ij} = u_i v_j \lambda^k + O(\mu^k) \quad \text{for some } \mu < \lambda
\]
Example: \( A = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \). Check that \( A^k = 4^k \begin{pmatrix} 2/5 & 3/5 \\ 2/5 & 3/5 \end{pmatrix} + (-1)^k \begin{pmatrix} 3/5 & -3/5 \\ -2/5 & 2/5 \end{pmatrix} \).
Here \( u = (2,3), \quad v = (1), \quad \lambda = 4. \)

Interpretation of the corollary for Markov chains: If \( A \) is the transition matrix for an irreducible, aperiodic Markov chain, then \( \lambda = 1, \quad u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \) and \( u = \pi \) is the unique stationary probability vector. In this case we get that for any \( i, j \in [n], \)
\[
p^{(k)}_{ij} = (A^k)_{ij} \xrightarrow[k \to \infty]{} \pi_j \quad \text{exponentially fast as } k \to \infty.
\]
I.e., after a long time the chain has forgotten its initial state and the distribution is very close to stationary.

Proof of the corollary: Let \( B = \frac{1}{\lambda} A. \) We have \( u B = u, \quad B v = v. \) If \( w \) is in one of the generalized eigenspaces \( \ker(A - \lambda I)^n \) for some eigenvalue \( \alpha \neq \lambda, \) then \( |\alpha| < \lambda, \) so
\[
B^k w = \frac{1}{\lambda^k} A^k w \xrightarrow[k \to \infty]{} 0 \quad \text{exponentially fast as } k \to \infty.
\]
(\( \|A^k \| \) is bounded by \( \kappa^k \) times a polynomial factor - exercise).
The same holds for row generalized eigenvectors. Therefore \( B^k \) converges to the unique matrix that satisfies \( B^k u = u \) and \( B^k w = 0 \) for any other generalized eigenvectors \( w, w'. \)
The matrix $N = uu$ is such a matrix: $uN = uu(vu) = (uv)u = u$, $Nv = (wu)v = (uv) = v$. If $w$ is as before a generalized eigenvalue for an eigenvalue $\lambda$, i.e., $\lambda v = \lambda w$, then

$$Mw = vuuw = vu(b^ku)w = vu(B^kw) \xrightarrow{k\to\infty} 0, \text{ and similarly}$$

$$w'M = w'uuw = (w'B^ku)u = (w'B^kw)u \xrightarrow{k\to\infty} 0 \text{ for a generalized row eigenvector } w'.$$

\[\square\]

2. **Finite state Markov chains**

Let $\Sigma$ be a finite set. A Markov chain with state space $\Sigma$ is a sequence of random variables $X_0, X_1, X_2, \ldots$, taking values in $\Sigma$ such that

$$P(X_n = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) = p_{i_{n-1}j}$$

where $P_{ij}$ is a fixed matrix of numbers called the transition matrix of the chain. Since $p_{ij} \geq 0$ represent transition probabilities, we have $\sum_j p_{ij} = 1$ for all $i$ -- the sums of all rows are 1. A matrix of nonnegative numbers satisfying this property is called stochastic. Note that this means that $A(\cdot|\cdot) = (\cdot|\cdot)$, i.e.,

$$1 = (\cdot|\cdot)$$

is a right-eigenvector of $A$ with eigenvalue 1.

The vector $\pi = (P(X_0 = i))_{i \in \Sigma}$ is called the initial state distribution. The following lemma shows how to get the distribution of the $n$-th state $X_n$ from $\pi$. 

Lemma 2: \( (P(X_n=j))_{j \in \Sigma} = \pi A^n \) (matrix multiplication).

Proof: Induction:
\[
P(X_n=j) = \mathbb{E} P(X_n=j | X_{n-1}) = \sum_{i \in \Sigma} P(X_{n-1}=i) P(X_n=j | X_{n-1}=i)
\]
\[
= \sum_{i \in \Sigma} \pi_i (\pi A)_{j,i} = (\pi A^n)_{j,n}
\]

Denote \( p^{(n)}_{i,j} = (\pi A)^n_{j,i} \). \( p^{(n)}_{i,j} \) represents the probability for a chain (non-random) with transition matrix \( A \) started from state \( i \in \Sigma \) to be at state \( j \in \Sigma \) after \( n \) steps. For a general initial state distribution \( \pi \), we have
\[
P(X_n=j) = \sum_i \pi_i p^{(n)}_{i,j}
\]

The chain is called irreducible if for any \( i,j \in \Sigma \) there exists an \( n \geq 1 \) s.t. \( p^{(n)}_{i,j} > 0 \), i.e., if the transition matrix is irreducible.

The chain is called aperiodic if the transition matrix is irreducible with period 1. (equivalently, \( A \) is primitive).

A state distribution vector \( \pi = (\pi_i)_{i \in \Sigma} \) is called a stationary vector for the chain if \( \pi A = \pi \) (i.e., \( \pi \) is a left eigenvector with eigenvalue 1). A chain started with this state distribution will satisfy
\[
(P(X_n=j))_{j \in \Sigma} = (\pi A^n)_{j,n} = \pi_j \text{ for all } n \geq 0.
\]
That is, the variables \( X_0, X_1, X_2, \ldots \) will be identically distributed.
Furthermore, because of the Markov property, it is easy to see that 
\( X_0, X_1, X_2, \ldots \) is actually a stationary sequence.

Applying the Perron-Frobenius theory to the case of a Markov chain, we get the following result.

**Theorem 3:** Let \( A \) be the transition matrix for an irreducible Markov chain. There exists a unique stationary distribution vector \( \pi = (\pi(i))_{i \in \Sigma} \). If the chain is aperiodic, then we have

\[
P^{(n)}_{ij} \xrightarrow{n \to \infty} \pi_j \quad \text{exponentially fast as } n \to \infty.
\]

3. **Recurrence and ergodicity of Markov chains**

**Theorem 4:** Let \( X_0, X_1, \ldots \) be an irreducible Markov chain. Then all states \( i \in \Sigma \) are recurrent, i.e.,

\[
P(X_n = i \text{ i.o.}) = 1.
\]

**Theorem 5:** Let \( X_0, X_1, \ldots \) be a stationary irreducible Markov chain. Then the sequence \( (X_n)_{n=0}^{\infty} \) is ergodic.

**Proof:** Let \( \pi \) be the stationary vector. The probability \( P(E) = P_n(E) \)

of any event can be written as \( P_n(E) = \sum_{j \in \Sigma} \pi_j P_j(E) \), where

\( P_j(\cdot) \) represents the distribution measure of the chain \( j \) with the same transition matrix that is started from state \( j \in \Sigma \).

Thus we have

\[
P(A) = \sum_j \pi_j P_j(A) = \sum_j \pi_j h(j),
\]

where we denote \( h(j) = P_j(A) = E_j 1_A \).
Let $S$ denote the shift transformation on $\mathbb{R}_+^{\mathbb{N}_0}$, and denote $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. We have $1_A = 1_A \circ S^n$ a.s., so

$$E(1_A | \mathcal{F}_n) = E(1_A \circ S^n | \mathcal{F}_n) = \sum_{j \in \mathbb{S}} 1_{X_n = j} E(1_A \circ S^n | X_n = j)$$

$$= \sum_{j \in \mathbb{S}} 1_{X_n = j} E(1_A | X_n = j) = \sum_j h(j) 1_{X_n = j} = h(X_n).$$

By Lévy's 0-1 law we have $E(1_A | \mathcal{F}_n) \xrightarrow{n \to \infty} 1_A$ a.s. So we have shown that $h(X_n) \xrightarrow{n \to \infty} 1_A$ a.s. Since by Theorem 4 all states are recurrent, it follows that $h = \text{const}$ and $h(X_n) \xrightarrow{n \to \infty} \text{const.} = 1_A$ a.s., so $P(A) = 0$ or 1, which proves ergodicity.

Proof of Theorem 4: We assume the period is 1 - the general case is left as an exercise. In that case the transition matrix $A$ is primitive. Let $N_0$ be such that all entries of $A^{N_0}$ are positive. Let $c = \text{the minimal entry of } A^{N_0}$. Fix $i \in \mathbb{S}$ and consider the sequence of events $E_m = \{X_{N_0m} = i\}$. We have $P(E_m | E_{m-1}, \ldots, E_1) = E[P(E_m | X_1, \ldots, X_{N_0(m-1)}) | E_{m-1}, \ldots, E_1] > c$, so by a coupling argument, the number $S_m = \sum_{k=1}^m 1_E_k$ of occurrences of $E_1, \ldots, E_m$ stochastically dominates a Binom$(m, c)$ r.v.; in fact $(S_1, S_2, S_3, \ldots)$ simultaneously dominate the process $(\sum_{k=1}^m Y_k)_{m \geq 1}$ where $(Y_m)$ are i.i.d. Bernoulli$(c)$. It follows that $P(E_m \text{ i.o.}) = P(Y_k = 1 \text{ i.o.}) = 1$. 
\qed