Definition 1. A *transition* (or stochastic) *matrix* is a square matrix with non-negative entries whose rows add up to 1. Say a transition matrix is *positive* if all its entries are positive.

Theorem 1. Eigenvalues of a Transition Matrix.
Let W be a transition matrix. The eigenvalues (real or complex) of W have magnitude at most 1. In the special case where W is positive, the only eigenvalue of magnitude 1 is 1.

Proof. Let $W\vec{v} = \lambda\vec{v}$ and say the component v_m of $\vec{v}$ has largest magnitude. Then

$$|\lambda v_m| = \left|\sum_k W_{m,k} v_k\right| \leq \sum_k W_{m,k}|v_k| \leq \sum_k W_{m,k}|v_m| = |v_m|$$

and thus $|\lambda| \leq 1$. If equality holds, then $W_{m,k}v_k$ all point in the same direction in the complex plane, and $|v_k| = |v_m|$ for all k. For W with strictly positive entries, we have $v_k = v_m$ for all k, which from $W\vec{v} = \lambda\vec{v}$ gives $\lambda = 1$ (corresponding to the all-ones vector). ∎

Theorem 2. Jordan Form of a Transition Matrix.
Let Λ be a Jordan Form of a transition matrix $W = P\Lambda P^{-1}$. The Jordan blocks of all eigenvalues of W of magnitude 1 have size 1.

Proof. Note that $\Lambda^n = P^{-1}W^nP$ is bounded since W^n is bounded. So powers of Jordan blocks are bounded:

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots \\ 0 & \lambda & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}^n = (\lambda I + N)^n = \lambda^n I + \binom{n}{1}\lambda^{n-1}N + \cdots = \begin{bmatrix} \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \cdots \\ 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \qquad (*)$$

For $|\lambda| = 1$, this power is unbounded unless the Jordan block has size 1. ∎

Theorem 3. Convergence of a Transition Matrix.
If W^n is positive for all $n \geq N$ for some N, then W^n converges as $n \to \infty$.

Proof. W^n positive $\Rightarrow$ The only eigenvalue of W^n of magnitude 1 is 1. Let λ be an eigenvalue of W of magnitude 1. Then $\lambda^n = 1$ for all $n \geq N$ and hence $\lambda = 1$.

Therefore the only eigenvalue of W of magnitude 1 is 1. The Jordan block $[1]^n$ corresponding to 1 obviously converges to $[1]$. And the other Jordan blocks (corresponding to $|\lambda| < 1$) converge to the zero matrix as $n \to \infty$ (see $(*)$). Hence Λ^n converges $\Rightarrow$ W^n converges. ∎

Theorem 4. Existence of a Stationary Distribution.
A stationary distribution exists, i.e. $\vec{\pi} \geq 0$ ($\vec{\pi} \neq \vec{0}$) such that $\vec{\pi} = \vec{\pi}W$.

Proof. W has 1 as an eigenvalue, so $W^\top$ also has 1 as an eigenvalue. This guarantees a row vector $\vec{v}$ such that $\vec{v} = \vec{v}W$, but the problem is we don't know if $\vec{v} \geq 0$.

To tackle this, take any initial distribution $\vec{u}_0$. Then $\vec{u}_n := \vec{u}_0 W^n \geq 0$ for all n. The idea is to *average out* $\vec{u}_n$ as $n \to \infty$. Write $\vec{u}_0 = \sum_{i=0}^N c_i\vec{v}_i$ where $\vec{v}_i$ are the column vectors of P in $W = P\Lambda P^{-1}$. We can partition the set $\{\vec{v}_i\}$ into those with eigenvalue 1 (A), with eigenvalue with magnitude 1 but not 1 (B), and with eigenvalue with magnitude < 1 (C). Then

$$\vec{u}_n = \vec{u}_0 W^n = \sum_{v_i \in A} c_i\vec{v}_i + \sum_{v_i \in B} c_i\lambda_i^n\vec{v}_i + \sum_{v_i \in C} c_i\vec{v}_i W^n \geq 0$$

For large $n \geq N$, $\sum_{v_i \in C} c_i \vec{v}_i W^n$ vanishes (the Jordan blocks corresponding to those vectors vanish). We then average out the above equation over sufficiently large $n \geq N$. Then the λ_i^n terms will also vanish:

$$\lim_{T \to \infty} \frac{1}{T+1} \sum_{n=N}^{N+T} \lambda_i^n = \lambda^N \lim_{T \to \infty} \frac{1 - \lambda^T}{(T+1)(1-\lambda)} = 0.$$

Leaving us with $\sum_{v_i \in A} c_i \vec{v}_i \geq 0$, which is a valid $\vec{\pi} \geq 0$ with $\vec{\pi} = \vec{\pi} W$. It is not the zero vector because the average of any $\vec{u}_n$ over some range of n is still a distribution, i.e. its components still add up to 1. ■

From now on, let $\mathcal{X} = \{X_t : t \in \mathbb{N}\}$ be a time-homogenous Markov process, where $X_t \in \mathcal{D}$ for all $t \in \mathbb{N}$. Let $G_{\mathcal{X}}$ be its Markov chain, and W be its transition matrix.

Theorem 5. n-step Probabilities.
The i, j-th entry of W^n is the probability that state i ends up in state j with exactly n steps, i.e.

$$(W^n)_{i,j} = \mathbb{P}(X_n = j \mid X_0 = i) := p_{i,j}^{(n)}$$

Proof. We write $p_{i,j}^{(n)}$ recursively:

$$\mathbb{P}(X_n = j \mid X_0 = i) = \sum_{d \in \mathcal{D}} \mathbb{P}(X_n = j \mid X_{n-1} = d) \, \mathbb{P}(X_{n-1} = d \mid X_0 = i)$$

$$p_{i,j}^{(n)} = \sum_{d \in \mathcal{D}} W_{d,j} p_{i,d}^{(n-1)} = \sum_{d \in \mathcal{D}} p_{i,d}^{(n-1)} W_{d,j}$$

but this is exactly the formula when multiplying by the matrix M. *Note: This recursive relation is known as the* Chapman-Kolmogorov Equation. ■

Theorem 6. Strongly Connected ⇒ Unique Stationary Distribution.
If $G_{\mathcal{X}}$ is strongly connected, then, $G_{\mathcal{X}}$ has a unique stationary distribution.

Proof. Since $G_{\mathcal{X}}$ is strongly-connected, for any states i, j there exists an n such that state i can get to state j in n steps, and so $p_{i,j}^{(n)} > 0$. By Theorem 4, there exists a stationary distribution $\vec{\pi}$.

- All components of $\vec{\pi}$ are strictly positive: If some j-th component of $\vec{\pi}$ is zero, it stays zero forever, which isn't possible since some other state will get to state j eventually (strong connectedness).
- Assume there were two stationary distributions $\vec{x}$ and $\vec{y}$. Since they are strictly positive, we can write $\vec{x} \geq r\vec{y}$ with one of the components holding a tight equality (say $x_m = ry_m$). Then for any n, we have $\vec{x} = \vec{x} W^n$ and $\vec{y} = \vec{y} W^n$, so

  $$x_m = \sum_i x_i (W^n)_{i,m} \geq r \sum_i y_i (W^n)_{i,m} = r y_m = x_m,$$

  so $x_i (W^n)_{i,m} = r y_i (W^n)_{i,m} \implies x_i p_{i,m}^{(n)} = r y_i p_{i,m}^{(n)}$ for all i, n. Since we can always choose an n such that $p_{i,m}^{(n)} > 0$, we have $x_i = r y_i$ for all i. Hence $\vec{x} = r\vec{y}$, giving $\vec{x} = \vec{y}$. ■

Theorem 7. Fundamental Theorem of Markov Chains.
If $G_{\mathcal{X}}$ is strongly connected and aperiodic, then every random walk on $G_{\mathcal{X}}$ converges to the unique stationary distribution.

Proof. When $G_{\mathcal{X}}$ is strongly connected, any state i can get to any other state j eventually. When $G_{\mathcal{X}}$ is furthermore aperiodic, by the Chicken McNugget Theorem, there exists an N such that state i can get to state j in any $n \geq N$ steps. Therefore $W^n, W^{n+1}, \cdots$ are positive, so by Theorem 3, W^n converges to some W', and that given any initial distribution $\vec{\pi}$, eventually $W^n\vec{\pi}$ converges to $W'\vec{\pi}$. ■

Definition 2. A strongly connected component of a Markov Chain $G_{\mathcal{X}}$ is also called a *communicating class*. Classes that have an outward edge to other classes are called *transient classes*, while classes that don't are called *(positive) recurrent classes*. (For infinite Markov Chains, however, there is a distinction between positive recurrent and null recurrent.)

Theorem 8. Aperiodic ⇒ Convergence.
If $G_{\mathcal{X}}$ is aperiodic, every random walk on $G_{\mathcal{X}}$ converges to some stationary distribution.

Proof. Consider the communicating classes of $G_{\mathcal{X}}$. If a class is transient, we can expect that the total probabilities in that class will diminish to 0 when $n \to \infty$ as the probabilities *leak out* from the class. Therefore, eventually we can approximate a distribution purely on the recurrent classes. Each recurrent class is strongly connected and aperiodic, and hence any random walk on them converges. ■

Theorem 9. Exact Condition of Uniqueness.
The following are equivalent:

- $G_{\mathcal{X}}$ has a unique stationary distribution.
- $G_{\mathcal{X}}$ has exactly one recurrent class.

Proof. Firstly, all stationary distributions must have 0 on all states in the transient classes (probability leak out). Therefore any stationary distribution is specified by a stationary distribution on each of the recurrent classes.

If there is only one recurrent class, then by Theorem 6, the stationary distribution is unique. If there are more than one recurrent class, since the recurrent classes don't flow to each other, any weighted average of the stationary distributions among the recurrent classes gives a valid stationary distribution, so the stationary distribution is not unique. ■

Theorem 10. Exact Condition of Convergence.
The following are equivalent:

- All recurrent classes of $G_{\mathcal{X}}$ are aperiodic.
- Every random walk of $G_{\mathcal{X}}$ converges.

Proof. Again we just have to look at the recurrent classes due to probability leak outs. The forward direction is Theorem 8. On the other hand, if some recurrent class is periodic with period d, we first pick a random state i in the Markov Chain, then we implement a uniform distribution on all and only the states that are accessible from i by paths with lengths that are a multiple of d. Then $p_{ii}^{(n)} > 0$ if and only if n is a multiple of d, so this distribution does not converge. ■