THE GEOMETRY OF MARKOV CHAIN LIMIT THEOREMS

JAMES BERNHARD

Abstract. In this paper, we use a geometric viewpoint to prove several of the fundamental theorems on the convergence of Markov chains. In particular, we determine the long-term behavior of Markov chains geometrically both when they are irreducible and aperiodic and when they are not. By viewing the transition matrix of a Markov chain as a linear transformation from the standard simplex to itself, we simplify the traditional, matrix-based descriptions and proofs of the long-term behavior of Markov chains.

In this paper, we prove several of the fundamental theorems on the convergence of Markov chains from a geometric point of view. While these theorems are well known and understood, this point of view seems not to be. Most of the key ideas in this paper are contained in Pullman's 1965 article *The Geometry of Markov Chains* [1], but that article seems not to be widely known or cited, and it has some errors and omissions. Most of them are relatively minor, but some of them occur in key parts of the exposition, which detracts from the article's readability. Also, some of Pullman's original arguments can be simplified and extended.

Our contention is that the statements and proofs of these standard theorems on the convergence of Markov chains can be made considerably more accessible by reframing them in geometric terms. We hope that this paper will serve to clarify, elaborate on, extend, and make more widely known the geometric point of view of Markov chain convergence laid out by Pullman.

Most of the results that we use in this paper are not standard, but in a few cases we include small standard theorems from areas outside of probability theory to keep this paper as self-contained as possible.

1. The main results in traditional terminology

To describe the results that we will prove in this paper, we first recall some standard terminology. Suppose that we have a Markov chain with states $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and an $n \times n$ transition matrix T whose (i, j)-th entry gives the probability of transitioning from $\mathbf{e}_j$ to $\mathbf{e}_i$.[1] As the Markov chain transitions from state to state, we refer to each step in which it is in a state as a *stage*, starting with Stage 0.

Definition 1. *A* **convex combination** *is a linear combination whose coefficients are nonnegative and sum to 1. A* **distribution** *is a convex combination of states. A Markov chain's distribution in Stage 0 is its* **initial distribution**. *If a distribution*

Date: June 2012.

1991 *Mathematics Subject Classification.* Primary: 60J02 Secondary: 52B02 [2010 mathematics subject classification].

Key words and phrases. Markov chain, convergence, convex polytope, geometry.

[1]This is the transpose of what is traditionally termed the transition matrix. As we will soon discuss, we use this instead for notational convenience.

involves only a subset of states (the coefficients of the other states being 0), we say that the distribution is **within** *those states.*

We usually interpret a distribution of states as a probability distribution, in the sense that the coefficient of each state gives its probability.

In order to describe the main results, we need some more terminology from Markov chain theory as well.

Definition 2. *A state* $\mathbf{e}_i$ **leads to** *a state* $\mathbf{e}_j$*, written* $\mathbf{e}_i \to \mathbf{e}_j$*, if there is a nonzero transition probability from* $\mathbf{e}_i$ *to* $\mathbf{e}_j$ *in some positive number of stages. Two states* $\mathbf{e}_i$ *and* $\mathbf{e}_j$ **communicate** *if* $\mathbf{e}_i \to \mathbf{e}_j$ *and* $\mathbf{e}_j \to \mathbf{e}_i$*. A state* $\mathbf{e}_i$ *is* **ergodic** *if it has the property that if* $\mathbf{e}_i \to \mathbf{e}_j$*, then* $\mathbf{e}_j \to \mathbf{e}_i$*. A state is* **transient** *if it is not ergodic. A Markov chain is* **irreducible** *if each state communicates with all the states.*

Directly from this definition, we can check that communicating is an equivalence relation among ergodic states, which means that it divides the ergodic states into equivalence classes.

Definition 3. *Among the ergodic states, the equivalence classes of communication are called* **ergodic classes***.*

Ergodic classes are sometimes referred to in the literature as a communication classes.

One of the main theorems that we will prove in this paper is the following.

Theorem 1. *For any initial distribution* $\mathbf{x}$ *and any transient state* $\mathbf{e}_i$*:*

$$\lim_{k\to\infty} \mathrm{Prob}(\mathbf{e}_i \text{ in Stage } k \mid \text{initial distribution } \mathbf{x}) = 0.$$

In other words, transient states are aptly named. They "die out" in the long run, in the sense that the probability of being in one approaches 0 as the number of stages that have been run tends to infinity.

What happens to ergodic states in the long run is more difficult to describe without using geometric terminology, but before we prove the main results about this geometrically, we at least state them here in more traditional terms. For this, the following definition will be useful.

Definition 4. *A positive integer* q *is a* **return time** *for a state* $\mathbf{e}_i$ *if there is a nonzero probability of being in state* $\mathbf{e}_i$ *in* q *stages given that the current state is* $\mathbf{e}_i$*. The* **period** *of a state* $\mathbf{e}_i$ *is the greatest common divisor of its return times. If* $\mathbf{e}_i$ *has no return times, then its period is infinity. A Markov chain is* **aperiodic** *if the period of each of its states is 1.*

We will prove the following theorem about the periods of ergodic states.

Theorem 2. *If* $\mathbf{e}_i$ *is ergodic, then it has a finite period and all states in its ergodic class have the same period.*

We will also prove the following theorem describing the long-term behavior of ergodic states.

Theorem 3. *Each ergodic class whose states have period* r *can be divided into* r **ergodic subclasses** $S_1, S_2, \ldots, S_r$ *with the property that if the Markov chain's distribution in Stage* k *is within* S_j*, then the Markov chain's distribution in Stage* $k+1$ *is within* S_{j+1} *(with the convention that when* $j = r$ *this index is* 1 *instead of* $r+1$*).*

Also, if $\mathbf{x}$ *is an initial distribution within* S_j*, then*

$$\lim_{k\to\infty} (\text{distribution in Stage } kr \text{ given initial distribution } \mathbf{x})$$

exists, is within S_j*, and is independent of the initial distribution within* S_j*.*

That is, transitioning the Markov chain cycles through the ergodic subclasses repeatedly, so that the distribution will be within any given ergodic subclass every r stages. And no matter what the initial distribution within an ergodic subclass is, the sequence of distributions in multiples of r stages converges to the same limiting distribution within that ergodic subclass.

The terminology *ergodic subclass* is not standard, but the author is not aware of any standard terminology to replace it with.

One more term is useful to describe the results that we will prove.

Definition 5. *A distribution* $\mathbf{v}$ *such that* $T\mathbf{v} = \mathbf{v}$ *is called a* **stationary distribution**.

With this, we can now state the other two main results that we will prove. In the statement of this theorem, recall that T is the transition matrix of the Markov chain.

Theorem 4. *If a Markov chain is irreducible and aperiodic, then it has a unique stationary distribution* $\mathbf{v}$*. Furthermore, for any initial distribution* $\mathbf{x}$*:*

$$\lim_{k\to\infty} T^k\mathbf{x} = \mathbf{v}.$$

Theorem 5. *If* T^k *has all nonzero coefficients for some positive integer* k*, then it has a unique stationary distribution* $\mathbf{v}$*. Also, for any initial distribution* $\mathbf{x}$*:*

$$\lim_{k\to\infty} T^k\mathbf{x} = \mathbf{v}.$$

The paper's main results are awkward even to state without using geometric language, but we have done so here so that the reader can connect the geometric point of view that we will use with the more common terminology. We now begin to develop this geometric view.

2. A geometric approach

In brief, our approach to restating and proving the main results is to replace a Markov chain's matrix T of transition probabilities with the linear transformation $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ that it represents and to recognize the set of distributions in this context as the standard $(n-1)$-simplex $\Delta \subset \mathbb{R}^n$. Roughly speaking, the question then becomes: what does $\boldsymbol{T}^k(\Delta)$ look like as $k \to \infty$, and how does $\boldsymbol{T}$ act on the "limit" of these sets? We now elaborate on this.

Suppose that we have a Markov chain whose states are $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Let T be the $n \times n$ matrix whose i, j-th entry is the probability that the chain will transition from $\mathbf{e}_j$ to $\mathbf{e}_i$. Accordingly, the entries of T are all nonnegative, and the entries in each column sum to 1. These two conditions mean that T is a *(column)-stochastic* matrix.

As mentioned in an earlier footnote, the matrix T is the transpose of what is traditionally called the transition matrix of the Markov chain. But analyzing the properties of T acting on the left is equivalent to analyzing the properties of the traditional transition matrix acting on the right, with only minor changes of language. Rather than having to make these minor changes of language many times throughout this paper, we will simply work with T instead of the traditional transition matrix.

If we think of the states $\mathbf{e}_1, \ldots, \mathbf{e}_n$ as the standard basis vectors in $\mathbb{R}^n$, then distributions are convex combinations of these basis vectors.[2] This leads us to the following definitions.

Definition 6. *The* **convex hull** *of a non-empty set of points* $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ *is the set of convex combinations of the points in the set:*

$$\mathrm{Hull}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \{\alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k \mid \alpha_1, \ldots, \alpha_k \geq 0 \text{ and } \sum_{i=1}^{k} \alpha_i = 1\}.$$

A set is **convex** *if it contains the convex hulls of all its nonempty subsets. The* **standard $(n-1)$-simplex** Δ *is the convex hull of the standard basis vectors* $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$. *An element of* Δ *is called a* **distribution**.

Notice that this definition of a distribution coincides with the traditional definition given in the previous section. The language here is simply more geometric.

Now if a Markov chain has a distribution $\mathbf{x} \in \Delta$ in Stage k, then the Markov chain has distribution $T\mathbf{x}$ in Stage $k+1$. The property that the matrix T has nonnegative entries and columns that sum to 1 is equivalent to the property that $T(\Delta) \subset \Delta$. This may seem obvious, but it is worth emphasizing because it is the key to switching from a matrix point of view to a geometric one.

In this paper, we will turn our attention from the matrix T to the linear transformation $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ that it represents with respect to the standard basis:

$$\boldsymbol{T}(\mathbf{x}) = T\mathbf{x}.$$

It is immediate that $\boldsymbol{T}(\Delta) \subset \Delta$ for the linear transformation defined this way. This leads us to the following defintion.

Definition 7. *A linear transformation* $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ *is* **stochastic** *if*

$$\boldsymbol{T}(\Delta) \subset \Delta.$$

Note that by restricting the domain of $\boldsymbol{T}$ to Δ, we can think of a stochastic linear transformation as a linear operator on Δ, which would be an equivalent definition.

So far, we have shifted from the traditional perspective on Markov chains to a more geometric point of view. We would now like to understand the effect of applying a stochastic linear transformation to distributions repeatedly *ad infinitum*.

For this, we need to understand some things about convex polytopes.

3. Convex polytopes

We first give a basic background on convex polytopes, including the necessary definitions and terminology. We will then prove some results about convex polytopes that we will use later in describing the convergence of Markov chains.

[2] We could instead work in an abstract vector space whose vectors are defined to be linear combinations of the states $\mathbf{e}_1, \ldots, \mathbf{e}_n$, but this doesn't add anything to the picture, so we use the obvious isomorphism of this abstract vector space with $\mathbb{R}^n$ to allow us to work directly in $\mathbb{R}^n$.

3.1. **Basic background.** Some additional terminology will help to describe certain types of sets of distributions.

Definition 8. *A set $P \subset \mathbb{R}^n$ is called a* **(bounded) convex polytope** *if it is the convex hull of a nonempty finite set of points. A point $\mathbf{v} \in P$ is called a* **vertex** *if it is not in the convex hull of any nonempty subset of P that doesn't include $\mathbf{v}$. We denote the set of vertices of P by $\mathcal{V}(P)$. A convex polytope $Q \subset P$ is called a* **convex subpolytope** *if $\mathcal{V}(Q) \subset \mathcal{V}(P)$.*

From these definitions, it follows readily that any convex polytope is the convex hull of its vertices.

Note that for any set $P \subset \mathbb{R}^n$ and any linear transformation $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$,

$$\mathrm{Hull}(\boldsymbol{T}(P)) = \boldsymbol{T}(\mathrm{Hull}(P)).$$

In particular, this means that if P is a convex polytope, then so is $\boldsymbol{T}(P)$.

Also note that $\mathcal{V}(\Delta)$ is precisely the set of states for the Markov chain, so we can now give a precise geometric definition of a state.

Definition 9. *A* **state** *is an element of $\mathcal{V}(\Delta) = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.*

This is equivalent to the traditional definition of a state but is more geometric.

To talk about relationships among distributions, the following terminology is also useful.

Definition 10. *A finite set of points in $\mathbb{R}^n$ is* **affinely independent** *if the only linear combination of them with coefficients summing to 0 that equals the zero vector is the trivial linear combination (with all coefficients equal to 0). A convex polytope is called a* **simplex** *if its vertices form an affinely independent set.*

Note that the standard $(n-1)$-simplex is appropriately named, since it actually is a simplex by this definition.

A particularly important concept for our investigations here is the following. While this definition holds in arbitrary polytopes, we will use it in Δ throughout this paper, so we restrict our definition to such cases.

Definition 11. *The* **carrier** *of a subset of Δ is the smallest convex subpolytope of Δ containing that subset. We denote the carrier of a single-point set $\{\mathbf{x}\} \subset \Delta$ by $\mathcal{C}(\mathbf{x})$ and, more generally, the carrier of a set P by $\mathcal{C}(P)$.*

The main reason that carriers are so important here has to do with their interpretation in terms of Markov chains. To understand this, first note that the vertices of carriers are states, being vertices of Δ. More specifically, if $P \subset \Delta$, then $\mathcal{V}(\mathcal{C}(P))$ are the possible states for the Markov chain when its distribution is in P. Similarly, if the current distribution of a Markov chain is $\mathbf{x}$, then the Markov chain has nonzero probability of being in any state in $\mathcal{V}(\mathcal{C}(\mathbf{x}))$, and it has probability 0 of being in any state not in $\mathcal{V}(\mathcal{C}(\mathbf{x}))$. In other words:

$$\mathbf{x} = \sum_{k \,|\, \mathbf{e}_k \in \mathcal{V}(\mathcal{C}(\mathbf{x}))} \alpha_k \mathbf{e}_k \qquad \text{with all } \alpha_k > 0 \text{ and } \sum_k \alpha_k = 1.$$

Note that each α_k in the above expression is nonzero. This means that $\mathbf{x}$ is in the interior of $\mathcal{C}(\mathbf{x})$ unless $\mathbf{x}$ is itself a state (which would mean that $\mathcal{C}(\mathbf{x}) = \{\mathbf{x}\}$).

In Δ, we can simplify the question of whether or not a state $\mathbf{e}_i$ is in $\mathcal{C}(P)$ by exploiting the fact that the vertices $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of Δ form an orthonormal set. Because of this, if we write a point $\mathbf{x} \in \Delta$ relative to these vertices as

$$\mathbf{x} = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k,$$

then if the current distribution of a Markov chain is $\mathbf{x}$, the probabilities of the various states are given by:

$$\mathrm{Prob}(\mathbf{e}_k) = \alpha_k = \mathbf{x} \cdot \mathbf{e}_k$$

for $k = 1, 2, \ldots, n$.

For $P \subset \Delta$, this means that $\mathbf{e}_i \in \mathcal{C}(P)$ if and only if there exists a point $\mathbf{p} \in P$ such that $\mathbf{p} \cdot \mathbf{e}_i \neq 0$. Although the orthonormality of the states isn't necessary for our results since there are other ways to express α_k, we will use this approach at times to simplify the exposition, as in the proof of the next proposition.

3.2. **Some useful propositions and lemmas.** We now have all of the necessary concepts from the theory of convex polytopes, so we now proceed to prove some propositions and lemmas that will be useful in proving the main results about Markov chains.

Proposition 1. *Let $P \subset \Delta$ be a convex polytope (but not necessarily a convex subpolytope) with vertices $\mathcal{V}(P) = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. Then*

$$\mathcal{V}(\mathcal{C}(P)) = \bigcup_{k=1}^{m} \mathcal{V}(\mathcal{C}(\mathbf{v}_k)).$$

Proof. Since carriers are subpolytopes, all of the vertices in the above equation are states, so we can write them as $\mathbf{e}_i$ for some i.

We claim that $\bigcup_{k=1}^{m} \mathcal{V}(\mathcal{C}(\mathbf{v}_k)) \subset \mathcal{V}(\mathcal{C}(P))$. For this, suppose that

$$\mathbf{e}_i \in \bigcup_{k=1}^{m} \mathcal{V}(\mathcal{C}(\mathbf{v}_k)),$$

which means that $\mathbf{e}_i \in \mathcal{C}(\mathbf{v}_j)$ for some j. Since $\mathbf{v}_j \in P$, then $\mathcal{C}(\mathbf{v}_j) \subset \mathcal{C}(P)$, so $\mathbf{e}_i \in \mathcal{C}(P)$, and being a vertex of Δ, it must also be a vertex of the subpolytope $\mathcal{C}(P)$. This implies that $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(P))$, which proves the claim.

We also claim that $\mathcal{V}(\mathcal{C}(P)) \subset \bigcup_{k=1}^{m} \mathcal{V}(\mathcal{C}(\mathbf{v}_k))$. To show this, suppose that $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(P)) \subset \mathcal{C}(P)$. By the discussion preceding this proposition, this means that there exists a point $\mathbf{p} \in P$ such that $\mathbf{p} \cdot \mathbf{e}_i \neq 0$. Since P is a convex polytope, it is the convex hull of its vertices, so

$$\mathbf{p} = \sum_{k=1}^{m} \beta_k \mathbf{v}_k$$

for some $\beta_1, \ldots, \beta_m \in \mathbb{R}$. Since $\mathbf{p} \cdot \mathbf{e}_i \neq 0$, then

$$\sum_{k=1}^{m} \beta_k \mathbf{v}_k \cdot \mathbf{e}_i \neq 0,$$

which means that $\mathbf{v}_j \cdot \mathbf{e}_i \neq 0$ for some j, so again by the above discussion $\mathbf{e}_i \in \mathcal{C}(\mathbf{v}_j)$. Since $\mathbf{e}_i$ is a vertex of Δ, it must also be a vertex of the subpolytope $\mathcal{C}(\mathbf{v}_j)$. So

$\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(\mathbf{v}_j))$ for some j, which means that $\mathbf{e}_i \subset \bigcup_{k=1}^{m} \mathcal{V}(\mathcal{C}(\mathbf{v}_k))$, which proves the second claim and completes the proof of the proposition. □

One of our tasks will be to show that a particular convex polytope is a simplex. For this, the following lemma will be useful.

Lemma 1. *Let $P \subset \Delta$ be a convex polytope (but not necessarily a subpolytope) with vertices*

$$\mathcal{V}(P) = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}.$$

If $\mathcal{C}(\mathbf{v}_i) \cap \mathcal{C}(\mathbf{v}_j) = \emptyset$ whenever $i \neq j$, then P is a simplex.

Proof. Consider an affine combination of the vertices that equals the zero vector:

$$\sum_{j=1}^{n} \alpha_j \mathbf{v}_j = \mathbf{0}, \quad \text{where } \sum_{j=1}^{n} \alpha_j = 0.$$

From each carrier $\mathcal{C}(\mathbf{v}_j)$, choose a single state $\mathbf{e}_{i_j} \in \mathcal{C}(\mathbf{v}_j)$. Then for each $k = 1, \ldots, m$, taking the dot product of both sides of the above equation with $\mathbf{e}_{i_k}$ gives:

$$\sum_{j=1}^{n} \alpha_j \mathbf{v}_j \cdot \mathbf{e}_{i_k} = 0.$$

Since the states $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form an orthogonal set, then points in disjoint carriers are orthogonal. By assumption the carriers of distinct vertices are disjoint here, so since $\mathbf{e}_{i_k} \in \mathcal{C}(\mathbf{v}_k)$, then $\mathbf{v}_j \cdot \mathbf{e}_{i_k}$ equals 0 when $j \neq k$ and is nonzero when $j = k$. This means that

$$\alpha_k \mathbf{v}_k \cdot \mathbf{e}_{i_k} = 0,$$

so $\alpha_k = 0$. This holds for all $k = 1, \ldots, m$, so the vertices of P are affinely independent, so P is a simplex. □

We will have repeated occasion to use the following lemma in our geometric investigations.

Lemma 2. *Let $\boldsymbol{T}$ be a stochastic linear transformation. Then for any subset $P \subset \Delta$,*

$$\boldsymbol{T}(\mathcal{C}(P)) \subset \mathcal{C}(\boldsymbol{T}(P)).$$

Proof. We claim that if $\mathbf{e}_i \in \mathcal{C}(P)$, then $\boldsymbol{T}(\mathbf{e}_i) \in \mathcal{C}(\boldsymbol{T}(P))$. To prove this, suppose that $\mathbf{e}_i \in \mathcal{C}(P)$, which means that there exists a point $\mathbf{p} \in P$, such that

$$\mathbf{p} = \sum_{j=1}^{n} p_j \mathbf{e}_j \in P \text{ with } p_i > 0.$$

Now for any $\mathbf{e}_k \notin \mathcal{C}(\boldsymbol{T}(P))$, we have $\boldsymbol{T}(\mathbf{p}) \cdot \mathbf{e}_k = 0$, since $\boldsymbol{T}(\mathbf{p}) \in \boldsymbol{T}(P)$. But this means that

$$0 = \boldsymbol{T}(\mathbf{p}) \cdot \mathbf{e}_k = \sum_{j=1}^{n} p_j \boldsymbol{T}(\mathbf{e}_j) \cdot \mathbf{e}_k.$$

Since $\boldsymbol{T}$ is stochastic, then $\boldsymbol{T}(\mathbf{e}_j) \cdot \mathbf{e}_k \geq 0$ for all $j = 1, \ldots, n$. Since $p_i > 0$ also, the fact that this sum equals 0 tells us that $\boldsymbol{T}(\mathbf{e}_i) \cdot \mathbf{e}_k = 0$, so $\mathbf{e}_k \notin \mathcal{C}(\boldsymbol{T}(\mathbf{e}_i))$.

This means that if we expand $\boldsymbol{T}(\mathbf{e}_i)$ relative to the states $\mathbf{e}_1, \ldots, \mathbf{e}_n$, the only possibly nonzero coefficients are of states that are in $\mathcal{C}(\boldsymbol{T}(P))$. By the definition of the carrier, this proves that claim that $\boldsymbol{T}(\mathbf{e}_i) \subset \mathcal{C}(\boldsymbol{T}(P))$.

To complete the proof of the lemma, suppose that the vertices of $\mathcal{C}(P)$ are $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_m}$. Then

$$\boldsymbol{T}(\mathcal{C}(P)) = \boldsymbol{T}(\mathrm{Hull}(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_m})) = \mathrm{Hull}(\boldsymbol{T}(\mathbf{e}_{j_1}), \ldots, \boldsymbol{T}(\mathbf{e}_{j_m})).$$

By the above claim, $\boldsymbol{T}(\mathbf{e}_{j_1}), \ldots, \boldsymbol{T}(\mathbf{e}_{j_m}) \in \mathcal{C}(\boldsymbol{T}(P))$. But $\mathcal{C}(\boldsymbol{T}(P))$ is convex, so

$$\mathrm{Hull}(\boldsymbol{T}(\mathbf{e}_{j_1}), \ldots, \boldsymbol{T}(\mathbf{e}_{j_m})) \subset \mathcal{C}(\boldsymbol{T}(P)),$$

which proves the lemma. □

Some results about nested sequences of convex polytopes will be useful. For any nested sequence of convex polytopes $P_0 \supset P_1 \supset P_2 \supset \ldots$, we use the standard notation that $\bigcap_{k=0}^{\infty} P_k$ is the set of all points that are in P_k for all $k \in \mathbb{N}$. (We use $\mathbb{N}$ in this paper to denote the set of nonnegative integers and $\mathbb{Z}^+$ to denote the set of positive integers.)

Lemma 3. *Let $P_0, P_1, P_2, \ldots$ be convex polytopes with $P_0 \supset P_1 \supset P_2 \supset \ldots$, and let $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \ldots$ be a sequence with $\mathbf{p}_k \in P_k$ for all $k \in \mathbb{N}$. If $\{\mathbf{p}_{k_1}, \mathbf{p}_{k_2}, \mathbf{p}_{k_3}, \ldots\}$ is a subsequence that converges to $\mathbf{p}$, then*

$$\mathbf{p} \in \bigcap_{k=0}^{\infty} P_k.$$

Proof. Since each P_k is a convex polytope, it is compact. The intersection of compact sets is compact and so contains all its limit points. □

Corollary 1. *Let $P_0, P_1, P_2, \ldots$ be convex polytopes with $P_0 \supset P_1 \supset P_2 \supset \ldots$. Then $\bigcap_{k=0}^{\infty} P_k$ is nonempty.*

Proof. For each $k \in \mathbb{N}$, choose a point $\mathbf{p}_k \in P_k$. Since each P_k is compact, then so is $\bigcap_{k=0}^{\infty} P_k$, which means that the sequence $\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \ldots\}$ contains a convergent subsequence. By Lemma 3, the limit of this subsequence is in $\bigcap_{k=0}^{\infty} P_k$. □

Lemma 4. *Let P_0 be a convex polytope, and let $P_0, P_1, P_2, \ldots$ be convex polytopes with $P_0 \supset P_1 \supset P_2 \supset \ldots$. Also let*

$$P = \bigcap_{k=0}^{\infty} P_k.$$

Suppose that U is an open set containing P. Then there exists a $k_0 \in \mathbb{N}$ such that $P_k \subset U$ for all $k \geq k_0$.

Proof. We prove this by contradiction. Suppose that for every k, there exists a point $\mathbf{p}_k \in P_k$ that is not in U. Since $\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \ldots\} \subset P_0$ and P_0 is compact, then this sequence must have a convergent subsequence. Also, this sequence (including the convergent subsequence) is contained in the complement of U, which is closed. So the limit of the convergent subsequence must be in the complement of U. However, by Lemma 3, this subsequence converges to a point in P, which is a contradiction because $P \subset U$.

Therefore $P_{k_0} \subset U$ for some $k_0 \in \mathbb{N}$. Since $P_0 \supset P_1 \supset \ldots$, this proves the lemma. □

Lemma 5. *Let P_0 be a convex polytope, and let $P_0, P_1, P_2, \ldots$ be convex polytopes with $P_0 \supset P_1 \supset P_2 \supset \ldots$ and such that*

$$\bigcap_{k=0}^{\infty} P_k = \{\mathbf{p}\}.$$

If $\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots\}$ *is a sequence with* $\mathbf{p}_k \in P_k$ *for all* k*, then*

$$\lim_{k\to\infty} \mathbf{p}_k = \mathbf{p}.$$

Proof. For any $\varepsilon > 0$, let U be an open ball of radius $\varepsilon/2$ centered at $\mathbf{p}$. By Lemma 4, there exists a $k_0 \in \mathbb{N}$, such that $P_k \subset U$ for all integers $k \geq k_0$. This means that $\{\mathbf{p}_{k_0}, \mathbf{p}_{k_0+1}, \mathbf{p}_{k_0+2}, \dots\} \subset P_{k_0} \subset U$, so

$$|\mathbf{p}_i - \mathbf{p}_j| < \varepsilon \quad \text{whenever } i, j \geq k_0,$$

which means that this sequence is Cauchy. Since P_0 is a convex polytope, it is compact and hence complete, so the sequence converges. By Lemma 3, its limit is in $\bigcap_{k=0}^{\infty} P_k = \{\mathbf{p}\}$, which proves the lemma. □

With these concepts and lemmas, we can now proceed with our exploration of the geometry of Markov chains.

4. Description of the limit set

Throughout this section, we let $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a stochastic linear transformation, keeping in mind its interpretation in terms of Markov chains. The main geometric object to investigate in order to understand the convergence of Markov chains is the following, which we will use so much in this paper that we give it a name.

Definition 12. *Let* $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ *be a stochastic linear transformation. Its* **limit set** L *is defined as:*

$$L = \bigcap_{k=0}^{\infty} \boldsymbol{T}^k(\Delta).$$

By Corollary 1, L is nonempty.

The main results in this section are:

1. L is a simplex.
2. $\boldsymbol{T}$ permutes the vertices of L.
3. $\boldsymbol{T}$ moves the carrier of each vertex of L around among other carriers in the orbit of the vertex cyclically, collapsing each carrier to a point in the process.
4. The fixed point set of $\boldsymbol{T}$ is the convex hull of the barycenters of the orbits of the vertices of L.

We now prove these through a series of propositions and corollaries.

Proposition 2. *L is a convex polytope.*

Proof. Since $\boldsymbol{T}$ is a linear transformation, then for each $k \in \mathbb{N}$, the set $\boldsymbol{T}^k(\Delta)$ is the convex hull of

$$E_k = \{\boldsymbol{T}^k(\mathbf{e}_1), \boldsymbol{T}^k(\mathbf{e}_2), \dots, \boldsymbol{T}^k(\mathbf{e}_n)\},$$

so it is a convex polytope.

To show that L is a convex polytope, we find a set whose convex hull is L. The process for this is as follows. Let I_1 denote the sequence of nonnegative integers:

$$I_1 := \{0, 1, 2, 3, \dots\}.$$

The subscript in I_1 refers to the fact that I_1 will be the sequence-indexing set in the first step of the process.

For each $i \in I_1$, choose a point from E_i. This gives a sequence S_1 of points in Δ, indexed on I_1. Since Δ is compact, there is a subsequence S'_1 indexed on $I_2 \subset I_1$ that converges to some point $\mathbf{s}_1 \in \Delta$. By Lemma 3, $\mathbf{s}_1 \in L$. We will soon show that $\mathbf{s}_1$ is part of a finite set whose convex hull is L, but first we move to the next step in the process of finding that set.

If possible, for each $i \in I_2$ choose a point from E_i that is not in S'_1. If this is not possible, then the process is finished and we will show that we have found a finite set whose convex hull is L. If it is possible, then it gives a sequence S_2 in Δ indexed on I_2, which must have a subsequence S'_2 indexed on $I_3 \subset I_2$ that converges to some point $\mathbf{s}_2 \in L$.

The process continues: if possible, for each $i \in I_k$ choose a point from E_i that is not in S'_j for any $j < k$. If this is not possible, then the process is finished and we will show that we have found a finite set whose convex hull is L. If it is possible, then it gives a sequence S_k in Δ indexed on I_k, which must have a subsequence S'_k indexed on $I_{k+1} \subset I_k$ that converges to some point $\mathbf{s}_k \in L$, and we have found what we will show is another point in a finite set whose convex hull is L.

Repeat this process until it is no longer possible. The process will certainly stop after at most n steps because each E_k consists of only n points.[3]

We note for future reference that if $k > k_0$, then the number of elements in E_k is less than or equal to the number of elements in E_{k_0}. This is simply because if $\boldsymbol{T}^{k_0}(\mathbf{e}_i) = \boldsymbol{T}^{k_0}(\mathbf{e}_j)$, then $\boldsymbol{T}^k(\mathbf{e}_i) = \boldsymbol{T}^k(\mathbf{e}_j)$ for all $k > k_0$. In other words, once two points in the set have been mapped to the same point, they cannot be separated by further iterations of $\boldsymbol{T}$. Consequently, when all of the points in E_{k_0} have been included in convergent subsequences in the above process (so that the process ends), then all of the points in E_k for all $k > k_0$ have also been included in convergent subsequences.

We now show that the convex hull of the points $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_q$ found by the above process is L. Let

$$H = \mathrm{Hull}(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_q).$$

Since L is the intersection of convex sets, it is itself convex. This means that since $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_q \in L$, then $H \subset L$, meaning that if $\mathbf{x} \in H$, then $\mathbf{x} \in L$.

To complete the proof that the convex hull of $\mathbf{s}_1, \ldots, \mathbf{s}_q$ is L, we now show that if $\mathbf{x} \notin H$, then $\mathbf{x} \notin L$. For this, suppose $\mathbf{x} \in \Delta$ but $\mathbf{x} \notin H$. Since H is a convex polytope, it is closed, so there is a minimum distance $\mathrm{Dist}(\mathbf{x}, H) > 0$ between $\mathbf{x}$ and all points in H. Now suppose that in the above process for finding $\mathbf{s}_1, \ldots, \mathbf{s}_q$, the set E_k ran out of points. (This must be true for some k.) Then as discussed above, for all $j > k$ every point in E_j is in a sequence converging to some $\mathbf{s}_i \in L$. This means that for any $\varepsilon > 0$, there exists a k_0 such that for all $k > k_0$ the distance from every point in E_k to H is less than ε. (We can choose k_0 to be the largest index required among the t sequences converging to $\mathbf{s}_1, \ldots, \mathbf{s}_q$.) But if every point in E_k has distance to H that is less that ε, then the distance from every point in the convex hull of E_k (which is $\boldsymbol{T}^k(\Delta)$) to H is also less than ε.

Applying this to $\varepsilon = \mathrm{Dist}(\mathbf{x}, H)$, we find that whenever k is suitably large, the distance from any point in $\boldsymbol{T}^k(\Delta)$ to H is strictly less than $\mathrm{Dist}(\mathbf{x}, H)$. Consequently $\mathbf{x} \notin \boldsymbol{T}^k(\Delta)$, so $\mathbf{x} \notin L$.

[3] A statement of the necessity of this finiteness was omitted from Pullman's paper [P], but he includes it correctly in an analogous setting in a later paper.

We have now shown that $\mathbf{x} \in L$ if and only if $\mathbf{x} \in H$, so $L = H$. Since H is a convex polytope, this completes the proof that L is a convex polytope. □

Since the limit set L is a convex polytope, it is the convex hull of its vertices, so from now on we will focus on its vertices $\mathbf{v}_1, \ldots, \mathbf{v}_m$ rather than on the points $\mathbf{s}_1, \ldots, \mathbf{s}_q$ that we found in the proof of this proposition. Since we will use these vertices so often, we introduce some notation for them.

Definition 13. *We use m to denote the number of vertices in L, which we will denote by*

$$\mathcal{V}(L) = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}.$$

We now investigate the effect of applying $\boldsymbol{T}$ to L.

Proposition 3. $\boldsymbol{T}(L) = L$.

Proof. If $\mathbf{x} \in L$, then $\mathbf{x} \in \boldsymbol{T}^k(\Delta)$ for all $k \geq 0$, which means that $\boldsymbol{T}(\mathbf{x}) \in \boldsymbol{T}^{k+1}(\Delta) \subset \boldsymbol{T}^k(\Delta)$ for all $k \geq 0$, so $\boldsymbol{T}(\mathbf{x}) \in L$. Therefore $\boldsymbol{T}(L) \subset L$, so to complete the proof of the proposition, we need only show that $L \subset \boldsymbol{T}(L)$.

For this, again let $\mathbf{x} \in L$. Then

$$\mathbf{x} \in \boldsymbol{T}^k(\Delta) = \boldsymbol{T}(\boldsymbol{T}^{k-1}(\Delta)) \quad \text{for all } k \geq 1.$$

This means that for each $k \geq 1$, there exists a point $\mathbf{p}_k \in \boldsymbol{T}^{k-1}(\Delta) \subset \Delta$ such that $\boldsymbol{T}(\mathbf{p}_k) = \mathbf{x}$. Since Δ is compact, the sequence $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots\}$ has a subsequence that converges to a point $\mathbf{p} \in \Delta$. By Lemma 3, $\mathbf{p} \in L$. Because $\boldsymbol{T}$ is a linear transformation, it is continuous, so

$$\boldsymbol{T}(\mathbf{p}) = \boldsymbol{T}(\lim_{k\to\infty} \mathbf{p}_k) = \lim_{k\to\infty} \boldsymbol{T}(\mathbf{p}_k) = \lim_{k\to\infty} \mathbf{x} = \mathbf{x}.$$

This shows that any point $\mathbf{x} \in L$ can be written as $\boldsymbol{T}(\mathbf{p})$ for some $\mathbf{p} \in L$, so $L \subset \boldsymbol{T}(L)$, which completes the proof. □

Proposition 4. $\boldsymbol{T}$ *permutes* $\mathcal{V}(L)$.

Proof. Denote the span of $\mathcal{V}(L)$ in $\mathbb{R}^n$ by

$$W := \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m),$$

from which it is immediate that $L = \operatorname{Hull}(\mathbf{v}_1, \ldots, \mathbf{v}_m) \subset W$.

Proposition 3 implies that $\boldsymbol{T}(\mathbf{v}_i) \in L \subset W$ for all i, so $\boldsymbol{T}(W) \subset W$. By the same proposition, for each i there exists an $\mathbf{x}_i \in L \subset W$ such that $\mathbf{v}_i = \boldsymbol{T}(\mathbf{x}_i)$, so $W \subset \boldsymbol{T}(W)$. Putting these together, we have that $\boldsymbol{T}(W) = W$.

This means that if we let $\boldsymbol{T}_{|W}$ denote the linear transformation $\boldsymbol{T}$ with its domain restricted by W, then $\boldsymbol{T}_{|W} : W \to W$ is invertible (since it is onto and therefore one-to-one by the rank-nullity theorem). Also, by Proposition 3 we know that $\boldsymbol{T}(L) = L$, so restricting the domain of $\boldsymbol{T}_{|W}$ to $L \subset W$ we find that $\boldsymbol{T}_{|L} : L \to L$ is also invertible.

Now if $\boldsymbol{T}(\mathbf{v}_i) = \mathbf{x} \in L$, then ${\boldsymbol{T}_{|L}}^{-1}(\mathbf{x}) = \mathbf{v}_i$. If $\mathbf{x}$ is not a vertex of L, then $\mathbf{x}$ lies in the interior of a line segment of points contained in L, which means that ${\boldsymbol{T}_{|L}}^{-1}$ of that line segment is a line segment in L containing $\mathbf{v}_i$ in its interior, which would contradict the fact that $\mathbf{v}_i$ is a vertex. Therefore $\mathbf{x}$ must be a vertex of L.

This tells us that $\boldsymbol{T}(\mathcal{V}(L)) \subset \mathcal{V}(L)$. Since $\boldsymbol{T}$ is invertible on all of W including $\mathcal{V}(L)$, we have that $\boldsymbol{T}_{|\mathcal{V}(L)} : \mathcal{V}(L) \to \mathcal{V}(L)$ is invertible, meaning that $\boldsymbol{T}$ permutes $\mathcal{V}(L)$. □

Because $\boldsymbol{T}$ permutes $\mathcal{V}(L)$, the general theory of permutations and group actions gives us additional information. For one thing, if $\mathbf{v} \in \mathcal{V}(L)$, then the sequence $\mathbf{v}, \boldsymbol{T}(\mathbf{v}), \boldsymbol{T}^2(\mathbf{v}), \ldots$ has only finitely many distinct terms, which are permuted cyclically by $\boldsymbol{T}$ with some finite order r. This order r has the properties that it is the smallest positive integer satisfying $\boldsymbol{T}^r(\mathbf{v}) = \mathbf{v}$, and that $\boldsymbol{T}^q(\mathbf{v}) = \mathbf{v}$ if and only if q is a multiple of r. The considerations lead us to the following definition.

Definition 14. *Let $\mathbf{v} \in \mathcal{V}(L)$. Then the* **orbit** *of $\mathbf{v}$ (under $\boldsymbol{T}$) is defined as:*

$$\mathcal{O}(\mathbf{v}) = \{\boldsymbol{T}^k(\mathbf{v}) \mid k \in \mathbb{N}\}.$$

The number of distinct points in $\mathcal{O}(\mathbf{v})$ is called the **order** *of $\mathbf{v}$, denoted by* $\mathrm{Order}(\mathbf{v})$.

Since $\boldsymbol{T}$ permutes the vertices of L, the order of any vertex of L is finite, and vertices in the same orbit have the same order since $\boldsymbol{T}$ permutes each orbit cyclically. Also, the set of vertices of L is partitioned into orbits: each vertex lies in exactly one orbit. This description of how $\boldsymbol{T}$ affects the vertices of L will be important throughout this paper.

We can now complete the proof that L is not just a convex polytope, but a simplex. By Lemma 1, to show that it is a simplex we need only show that the carriers of its vertices are disjoint. This is a corollary of the next proposition, which is also interesting in its own right.

Proposition 5. *If $\mathbf{v} \in \mathcal{V}(L)$, then*

$$\mathcal{C}(\mathbf{v}) \cap L = \{\mathbf{v}\}.$$

Proof. Since $\mathbf{v} \in \mathcal{V}(L)$, then $\mathbf{v} \in L$. Also, $\mathbf{v} \in \mathcal{C}(\mathbf{v})$ by the definition of a carrier, so $\mathbf{v} \in \mathcal{C}(\mathbf{v}) \cap L$.

Now if $\mathcal{C}(\mathbf{v}) = \{\mathbf{v}\}$, then the proposition follows trivially, so assume that $\mathcal{C}(\mathbf{v}) \neq \{\mathbf{v}\}$. In this case, $\mathbf{v} \in \mathrm{Interior}(\mathcal{C}(\mathbf{v}))$ as discussed in the previous section.

Suppose that $\mathbf{x} \in \mathcal{C}(\mathbf{v}) \cap L$. In order to arrive at a contradiction, assume that $\mathbf{x} \neq \mathbf{v}$. In this case, $\mathbf{v}$ and $\mathbf{x}$ determine a line, which can be parametrized as

$$(1-t)\mathbf{v} + t\mathbf{x},$$

where $t \in \mathbb{R}$. Since $\mathcal{C}(\mathbf{v})$ is convex, we know that $(1-t)\mathbf{v}+t\mathbf{x} \in \mathcal{C}(\mathbf{v})$ when $t \in [0,1]$. However, since $\mathbf{v} \in \mathrm{Interior}(\mathcal{C}(\mathbf{v}))$, this line segment can be extended in $\mathcal{C}(\mathbf{v})$ past $\mathbf{v}$. We now use this to find a line segment in L containing $\mathbf{v}$ not as an endpoint.

Since $\mathbf{x} \in L$ then for each $k \in \mathbb{N}$ there exists a point $\mathbf{x}_k \in \Delta$ such that

$$\boldsymbol{T}^k(\mathbf{x}_k) = \mathbf{x}.$$

Also, since each $\mathbf{x}_k \in \Delta$ and Δ is compact, then

$$M := \sup_k |\mathbf{x}_k - \mathbf{v}|$$

exists.[4] Since $\mathbf{v} \in \mathrm{Interior}(\mathcal{C}(\mathbf{v}))$, then for suitably small $\delta > 0$, a ball of radius δ about $\mathbf{v}$ is contained in $\mathcal{C}(\mathbf{v})$. This means that

$$(1-t)\mathbf{v} + t\mathbf{x}_k \in \mathcal{C}(\mathbf{v}) \quad \text{if} \quad |t| \leq \frac{\delta}{|\mathbf{x}_k - \mathbf{v}|}.$$

[4]The role played by the existence of this supremum is a subtlety that seems to have been overlooked in [P], but it is crucial to the argument. For example, if $\boldsymbol{T}$ were a linear operator on $\mathbb{R}^n$ with eigenvalues all less than 1 instead of a linear operator on Δ, then this part of the proof would fail. For this proposition, it is important that $\boldsymbol{T}(\Delta) \subset \Delta$.

Applying $\boldsymbol{T}^k$, it follows that

$$\boldsymbol{T}^k((1-t)\mathbf{v}+t\mathbf{x}_k) = (1-t)\mathbf{v}+t\mathbf{x} \in \boldsymbol{T}^k(\mathcal{C}(\mathbf{v})) \quad \text{if} \quad |t| \leq \frac{\delta}{|\mathbf{x}_k - \mathbf{v}|}.$$

But $\delta/|\mathbf{x}_k - \mathbf{v}| \geq \delta/M$, so if $|t| \leq \delta/M$, then $|t| \leq \delta/|\mathbf{x}_k - \mathbf{v}|$. This means that

$$(1-t)\mathbf{v}+t\mathbf{x} \in \boldsymbol{T}^k(\mathcal{C}(\mathbf{v})) \quad \text{if} \quad |t| \leq \frac{\delta}{M}.$$

This condition is independent of k, so it holds for all k, meaning that

$$(1-t)\mathbf{v}+t\mathbf{x} \in \bigcap_{k=0}^{\infty} \boldsymbol{T}^k(\mathcal{C}(\mathbf{v})) \quad \text{if} \quad |t| \leq \frac{\delta}{M}.$$

But this gives us a line segment in L containing $\mathbf{v}$ not as an endpoint, which is a contradiction since $\mathbf{v}$ is a vertex of L. Therefore $\mathbf{x} = \mathbf{v}$, meaning that $L \cap \mathcal{C}(\mathbf{v}) = \{\mathbf{v}\}$. $\square$

The first corollary to this proposition, roughly speaking, asserts that $\boldsymbol{T}$ maps the carriers of the vertices in an orbit cyclically, gradually collapsing each one to a point in the process.

Corollary 2. *Let* $\mathbf{v} \in \mathcal{V}(L)$ *with* $\text{Order}(\mathbf{v}) = r$, *and let* $q \in \mathbb{N}$. *Then*

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{q+kr}(\mathcal{C}(\mathbf{v})) = \boldsymbol{T}^q(\mathbf{v}).$$

Proof. First we prove the result when $q = 0$. For this, note that by Lemma 2,

$$\boldsymbol{T}^r(\mathcal{C}(\mathbf{v})) \subset \mathcal{C}(\boldsymbol{T}^r(\mathbf{v})) = \mathcal{C}(\mathbf{v}),$$

since $\boldsymbol{T}^r(\mathbf{v}) = \mathbf{v}$ by assumption. This means that $\boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v})) \subset \mathcal{C}(\mathbf{v})$ for all k, so

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v})) \subset \mathcal{C}(\mathbf{v}).$$

But we also know that:

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v})) \subset \bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\Delta) \subset \bigcap_{k=0}^{\infty} \boldsymbol{T}^{k}(\Delta) = L.$$

Putting these two together, Proposition 5 tells us that

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v})) \subset \mathcal{C}(\mathbf{v}) \cap L = \{\mathbf{v}\}.$$

This proves containment in one direction. For the other direction, note that $\boldsymbol{T}^r(\mathbf{v}) = \mathbf{v}$, so $\mathbf{v} \in \boldsymbol{T}^{kr}(\mathbf{v}) \subset \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v}))$ for all k.

Putting both directions together, we have that

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v})) = \{\mathbf{v}\},$$

which proves the result when $q = 0$.

Applying $\boldsymbol{T}^q$ to both sides of this equation completes the proof of the full proposition:

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{q+kr}(\mathcal{C}(\mathbf{v})) = \boldsymbol{T}^q\left(\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v}))\right) = \{\boldsymbol{T}^q(\mathbf{v})\}. \quad \square$$

The following additional corollary to this proposition establishes that the carriers of the vertices of L are disjoint.

Corollary 3. *If $\mathbf{v}_i, \mathbf{v}_j \in \mathcal{V}(L)$ and $i \neq j$, then $\mathcal{C}(\mathbf{v}_i) \cap \mathcal{C}(\mathbf{v}_j) = \emptyset$.*

Proof. Suppose that $\mathbf{x} \in \mathcal{C}(\mathbf{v}_i) \cap \mathcal{C}(\mathbf{v}_j)$. Let r_i, r_j be the orders of $\mathbf{v}_i, \mathbf{v}_j$ under the permutation $\boldsymbol{T}$, and let r be the least common multiple of r_i and r_j. Since r is a multiple of both r_i and r_j, then

$$\mathcal{C}(\mathbf{v}_i) \supset \boldsymbol{T}^r(\mathcal{C}(\mathbf{v}_i)) \supset \boldsymbol{T}^{2r}(\mathcal{C}(\mathbf{v}_i)) \supset \dots,$$

and similarly for $\mathbf{v}_j$.

Now $\boldsymbol{T}^{kr}(\mathbf{x}) \in \mathcal{C}(\mathbf{v}_i)$ for all k, and by Corollary 2:

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr}(\mathcal{C}(\mathbf{v}_i)) = \{\mathbf{v}_i\},$$

so Lemma 5 tells us that

$$\lim_{k\to\infty} \boldsymbol{T}^{kr}(\mathbf{x}) = \mathbf{v}_i.$$

By the same argument, replacing $\mathbf{v}_i$ with $\mathbf{v}_j$ throughout:

$$\lim_{k\to\infty} \boldsymbol{T}^{kr}(\mathbf{x}) = \mathbf{v}_j.$$

Therefore $\mathbf{v}_i = \mathbf{v}_j$, meaning that $i = j$, which is a contradiction. Therefore $\mathcal{C}(\mathbf{v}_i) \cap \mathcal{C}(\mathbf{v}_j) = \emptyset$. □

This has a couple of interesting corollaries.

Corollary 4. *Each state in $\mathcal{V}(\mathcal{C}(L))$ is in the carrier of exactly one vertex in $\mathcal{V}(L)$.*

Proof. By Proposition 1, we know that each state is in the carrier of some vertex in $\mathcal{V}(L)$, and Corollary 3 tells us that each state is in the carrier of at most one vertex in $\mathcal{V}(L)$, which proves this corollary. □

We will use the relationship in this corollary so often that we give it a name and introduce some notation for it.

Definition 15. *For any state $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(L))$, the unique vertex $\mathbf{v} \in \mathcal{V}(L)$ such that $\mathbf{e}_i \in \mathcal{C}(\mathbf{v})$ is called the* **limit vertex** *of $\mathbf{e}_i$, denoted by $v(\mathbf{e}_i)$.*

Some other concepts that we will use in describing the long-term behavior of Markov chains geometrically are the following.

Definition 16. *Let $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(L))$. The* **limit vertex class** *of $\mathbf{e}_i$ is defined as the set:*

$$\{\mathbf{e}_j \in \mathcal{V}(\mathcal{C}(L)) \mid v(\mathbf{e}_j) = v(\mathbf{e}_i)\}.$$

The **orbit class** *of $\mathbf{e}_i$ is defined as the set:*

$$\{\mathbf{e}_j \in \mathcal{V}(\mathcal{C}(L)) \mid v(\mathbf{e}_j) \in \mathcal{O}(v(\mathbf{e}_i))\}.$$

In other words, two states in the carrier of L are in the same limit vertex class if their limit vertices are the same. They are in the same orbit class if their limit vertices are in the same orbit. From these definitions, it is immediate that $\mathcal{V}(\mathcal{C}(L))$ is partitioned into orbit classes, each of which is in turn partitioned into limit vertex classes. We will use these definitions and concepts later in the paper.

But first we turn our attention to another of the main results in this section, which is implied by Corollary 3.

Corollary 5. *L is a simplex.*

Proof. This follows from Corollary 3 and Lemma 1. □

Also before we examine the limiting behavior of Markov chains, we describe the structure of the *fixed point set* of $\boldsymbol{T}$, which is the set of distributions $\mathbf{x} \in \Delta$ satisfying $\boldsymbol{T}(\mathbf{x}) = \mathbf{x}$. In traditional terms, fixed points of $\boldsymbol{T}$ are stationary distributions of the Markov chain that $\boldsymbol{T}$ represents.

Proposition 6. *The fixed point set of $\boldsymbol{T}$ is equal to the convex hull of the barycenters of the orbits of the vertices of L.*

Proof. Let $\mathbf{x}$ be the barycenter of the orbit of a vertex $\mathbf{v} \in \mathcal{V}(L)$ with $\mathrm{Order}(\mathbf{v}) = r$, so:

$$\mathbf{x} = \frac{1}{r}\sum_{k=0}^{r-1} \boldsymbol{T}^k(\mathbf{v}).$$

Since $\boldsymbol{T}^r(\mathbf{v}) = \boldsymbol{T}^0(\mathbf{v}) = \mathbf{v}$, then

$$\boldsymbol{T}(\mathbf{x}) = \frac{1}{r}\sum_{k=1}^{r} \boldsymbol{T}^k(\mathbf{v}) = \frac{1}{r}\sum_{k=0}^{r-1} \boldsymbol{T}^k(\mathbf{v}) = \mathbf{x},$$

so $\boldsymbol{T}$ fixes the barycenters of the orbits of the vertices of L. By linearity, $\boldsymbol{T}$ also fixes their convex hull, so the convex hull of the barycenters of the orbits is contained in the fixed point set of $\boldsymbol{T}$.

Now suppose $\mathbf{x} \in \Delta$ is an arbitrary point with $\boldsymbol{T}(\mathbf{x}) = \mathbf{x}$. Since $\boldsymbol{T}(\mathbf{x}) = \mathbf{x}$, then $\mathbf{x} \in L$, so $\mathbf{x}$ can be written as a convex combination of the vertices of L. Suppose that we choose one vertex $\mathbf{v}_{i_j} \in \mathcal{V}(L)$ from each of the t orbits, and that $\mathrm{Order}(\mathbf{v}_{i_j}) = r_{i_j}$. Writing $\mathbf{x}$ as a convex combination of the vertices, we have:

$$\mathbf{x} = \sum_{j=1}^{t} \sum_{k=0}^{r_{i_j}-1} \alpha_{j,k} \boldsymbol{T}^k(\mathbf{v}_{i_j}),$$

where each $\alpha_{jk} \geq 0$ and $\sum_{j,k} \alpha_{j,k} = 1$. Since $\boldsymbol{T}(\mathbf{x}) = \mathbf{x}$, we have:

$$\begin{aligned}
\sum_{j=1}^{t} \sum_{k=0}^{r_{i_j}-1} \alpha_{j,k} \boldsymbol{T}^k(\mathbf{v}_{i_j}) &= \mathbf{x} \\
&= \boldsymbol{T}(\mathbf{x}) \\
&= \sum_{j=1}^{t} \sum_{k=0}^{r_{i_j}-1} \alpha_{j,k} \boldsymbol{T}^{k+1}(\mathbf{v}_{i_j}) \\
&= \sum_{j=1}^{t} \sum_{k=0}^{r_{i_j}-1} \alpha_{j,k-1} \boldsymbol{T}^{k}(\mathbf{v}_{i_j}),
\end{aligned}$$

with the convention that $\alpha_{j,-1}$ denotes $\alpha_{j,r_{i_j}-1}$.

Since L is a simplex, then its vertices are affinely independent, so we can set corresponding coefficients in the above sums equal to each other:

$$\alpha_{j,k-1} = \alpha_{j,k} \qquad \text{for all } k = 0, 1, \ldots, r_{i_j} - 1.$$

Since $\alpha_{j,k}$ is independent of k, we can denote it simply by α_j for all k. This means that

$$\mathbf{x} = \sum_{j=1}^{t} r_{i_j}\alpha_j\Big(\frac{1}{r_{i_j}}\sum_{k=0}^{r_{i_j}-1}\boldsymbol{T}^k(\mathbf{v}_{i_j})\Big),$$

which is a convex combination of the barycenters of the orbits of the vertices, so the fixed point set of $\boldsymbol{T}$ is contained in the convex hull of the barycenters of the orbits of the vertices, which proves the proposition. □

5. Ergodic states and periods

In this section we will show:

(1) How to characterize ergodic (and therefore transient) states geometrically.
(2) That transient states "die out".
(3) What happens to ergodic states in the limit as the number of stages that the Markov chain is run tends to infinity.

First we translate from the traditional characterization of ergodic states to a more geometric one. In geometric terms, $\mathbf{e}_i \to \mathbf{e}_j$ means that

$$\mathbf{e}_j \in \bigcup_{k=1}^{\infty} \mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i)),$$

or equivalently that there exists a $k \in \mathbb{Z}^+$, such that $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i))$. (And of course, $\mathbf{e}_j$ is in a carrier if and only if it is a vertex of that carrier, since it is a vertex of Δ.) With this in mind, we prove the following proposition.

Proposition 7. *For any state $\mathbf{e}_i \in \mathcal{V}(\Delta)$, there exists some state $\mathbf{e}_j \in \mathcal{C}(L)$ such that $\mathbf{e}_i \to \mathbf{e}_j$.*

Proof. From a geometric point of view, $\mathbf{e}_i \to \mathbf{e}_j$ if and only if $\mathbf{e}_j \in \bigcup_{k=1}^{\infty}\mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i))$. To show this, we first claim that if $\mathbf{e}_j \in \mathcal{C}(\bigcup_{k=1}^{\infty}\boldsymbol{T}^k(\mathbf{e}_i))$, then $\mathbf{e}_j \in \bigcup_{k=1}^{\infty}\mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i))$. To prove this claim, suppose that $\mathbf{e}_j \in \mathcal{C}(\bigcup_{k=1}^{\infty}\boldsymbol{T}^k(\mathbf{e}_i))$. By definition, there is some $\mathbf{x} \in \bigcup_{k=1}^{\infty}\boldsymbol{T}^k(\mathbf{e}_i)$ for which

$$\mathbf{x} = \alpha\mathbf{e}_j + \text{terms involving other states}$$

with $\alpha > 0$. But since $\mathbf{x} \in \bigcup_{k=1}^{\infty}\boldsymbol{T}^k(\mathbf{e}_i)$, then by definition $\mathbf{x} = \boldsymbol{T}^q(\mathbf{e}_i)$ for some $q \in \mathbb{Z}^+$. This means that

$$\boldsymbol{T}^q(\mathbf{e}_i) = \alpha\mathbf{e}_j + \text{terms involving other states},$$

which means that $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i))$, so $\mathbf{e}_j \in \bigcup_{k=1}^{\infty}\mathcal{C}(\boldsymbol{T}^k(\mathbf{e}_i))$, as claimed.

For brevity of notation, define

$$C = \mathcal{C}(\bigcup_{k=1}^{\infty}\boldsymbol{T}^k(\mathbf{e}_i)).$$

By the above claim, to prove the proposition we need only show that there is some $\mathbf{e}_j \in \mathcal{C}(L)$ such that $\mathbf{e}_j \in C$. Lemma 2 tells us that

$$\boldsymbol{T}(C) \subset \mathcal{C}(\bigcup_{k=1}^{\infty}\boldsymbol{T}^{k+1}(\mathbf{e}_i)) \subset C,$$

so $C \supset \boldsymbol{T}(C) \supset \boldsymbol{T}^2(C) \supset \dots$, which means that $\bigcap_{k=0}^{\infty} \boldsymbol{T}^k(C)$ is nonempty by Corollary 1, and by its construction, it is a subset of C. Also, since $C \subset \Delta$, we have that

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^k(C) \subset \bigcap_{k=0}^{\infty} \boldsymbol{T}^k(\Delta) = L.$$

So $C \cap L$ contains $\bigcap_{k=0}^{\infty} \boldsymbol{T}^k(C)$, which means that C and L are not disjoint.

Since C and L have at least one point in common, their carriers have at least one state $\mathbf{e}_j \in \Delta$ in common:

$$\mathbf{e}_j \in \mathcal{C}(C) \cap \mathcal{C}(L).$$

But C is itself already a carrier, so $\mathcal{C}(C) = C$. This means that $\mathbf{e}_j \in \mathcal{C}(L)$ and $\mathbf{e}_j \in C$, which proves the proposition. □

We now prove another useful lemma.

Lemma 6. *Let $\mathbf{v} \in \mathcal{V}(L)$, and let $q \in \mathbb{N}$. Then there is a $k_0 \in \mathbb{N}$ for which*

$$\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{x})) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v})) \quad \text{for all } \mathbf{x} \in \mathcal{C}(\mathbf{v})$$

for all $k > k_0$.

Proof. For any $k \in \mathbb{N}$, we know that

$$\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathcal{C}(\mathbf{v}))) \subset \mathcal{C}(\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{v}))) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v})),$$

which means that for any $\mathbf{x} \in \mathcal{C}(\mathbf{v})$

$$\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{x})) \subset \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$$

for any $k \in \mathbb{N}$. We now prove that equality holds for suitably large k. If $\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ is a single point, then the lemma holds trivially, so assume that $\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ contains more than one point. In this case, $\boldsymbol{T}^q(\mathbf{v}) \in \text{Interior}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v})))$.

By Corollary 2, we know that

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr+q}(\mathcal{C}(\mathbf{v})) = \boldsymbol{T}^q(\mathbf{v}),$$

Since $\text{Interior}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v})))$ is an open set in $\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ containing this intersection, by Lemma 4 there exists a $k_0 \in \mathbb{N}$ such that $\boldsymbol{T}^{kr+q}(\mathcal{C}(\mathbf{v})) \subset \text{Interior}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v})))$ for all $k > k_0$. But this means that for any $\mathbf{x} \in \mathcal{C}(\mathbf{v})$,

$$\boldsymbol{T}^{kr+q}(\mathbf{x}) \subset \text{Interior}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v})))$$

for all $k > k_0$. If $\mathbf{x} \in \text{Interior}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v})))$ then $\mathbf{x}$ is in $\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ but not in any proper subpolytope of $\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$, so

$$\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{x})) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$$

for all $k > k_0$, which proves the lemma. □

We are now getting closer to a geometric characterization of ergodicity.

Proposition 8. *Let $\mathbf{e}_i \in \mathcal{C}(L)$ and $\mathbf{e}_j \in \mathcal{V}(\Delta)$. Then $\mathbf{e}_i \to \mathbf{e}_j$ if and only if $\mathbf{e}_j$ is in the orbit class of $\mathbf{e}_i$.*

Proof. For brevity, let $\mathbf{v} = v(\mathbf{e}_i)$. To prove the forward implication, suppose that $\mathbf{e}_i \to \mathbf{e}_j$, meaning that $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{e}_i))$ for some $q \in \mathbb{Z}^+$. By Lemma 2, we have:

$$\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{e}_i)) \subset \mathcal{C}(\boldsymbol{T}^q(\mathcal{C}(\mathbf{v}))) \subset \mathcal{C}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v})).$$

This means that $v(\mathbf{e}_j) = \boldsymbol{T}^q(\mathbf{v})$, so $\mathbf{e}_j$ is in the orbit class of $\mathbf{e}_i$, which proves the forward implication in the proposition.

For the other direction, suppose instead that $\mathbf{e}_j$ is in the orbit class of $\mathbf{e}_i$, which means that $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ for some $q \in \mathbb{N}$. If the order of $\mathbf{e}_i$ is r, then by Lemma 6, there is a $k_0 \in \mathbb{N}$ such that

$$\mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{e}_i)) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$$

for all integers $k > k_0$. Since $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$, this means that $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^{kr+q}(\mathbf{e}_i))$ for some k, so $\mathbf{e}_i \to \mathbf{e}_j$. □

We now give a geometric characterization of ergodicity in the following corollary.

Corollary 6. *A state $\mathbf{e}_i \in \mathcal{V}(\Delta)$ is ergodic if and only if $\mathbf{e}_i \in \mathcal{C}(L)$.*

Proof. Since $\mathbf{e}_i \in \mathcal{V}(\Delta)$, then $\mathbf{e}_i \in \mathcal{C}(L)$ if and only if $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(L))$. But by Corollary 4, we know that $\mathbf{e}_i \in \mathcal{V}(\mathcal{C}(L))$ if and only if $\mathbf{e}_i \in \mathcal{C}(\mathbf{v})$ for some vertex $\mathbf{v} \in \mathcal{V}(L)$.

To show that this condition is satisfied, suppose that $\mathbf{e}_i$ is not in the carrier of any vertex of L. Then by Proposition 7, there is some $\mathbf{v} \in \mathcal{V}(L)$ such that $\mathbf{e}_j \in \mathcal{C}(\mathbf{v})$ and $\mathbf{e}_i \to \mathbf{e}_j$. But since $\mathbf{e}_i$ is not in the carrier of any vertex of L, then $\mathbf{e}_i \notin \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ for any $q \in \mathbb{Z}^+$. By Proposition 8, it follows that $\mathbf{e}_j \not\to \mathbf{e}_i$. Since $\mathbf{e}_i \to \mathbf{e}_j$ but $\mathbf{e}_j \not\to \mathbf{e}_i$, then $\mathbf{e}_i$ is not ergodic.

For the other direction, suppose instead that there is some $\mathbf{v} \in \mathcal{V}(L)$ for which $\mathbf{e}_i \in \mathcal{C}(\mathbf{v})$. If $\mathbf{e}_i \to \mathbf{e}_j$, then by Proposition 8, $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ for some $q \in \mathbb{Z}^+$. If r is the order of $\mathbf{v}$, then choose a $q' \in \mathbb{Z}^+$ such that r divides $q + q'$, so that $\boldsymbol{T}^{q+q'}(\mathbf{v}) = \mathbf{v}$. This means that $\mathbf{e}_i \in \mathcal{C}(\mathbf{v}) = \mathcal{C}(\boldsymbol{T}^{q'}(\boldsymbol{T}^q(\mathbf{v})))$, so by Proposition 8 (in the other direction), $\mathbf{e}_j \to \mathbf{e}_i$. Therefore $\mathbf{e}_i$ is ergodic. □

We are now almost ready to describe ergodic classes and subclasses geometrically, but first we do the same for the period of an ergodic state.

Theorem 6. *For any ergodic state $\mathbf{e}_i \in \mathcal{V}(L)$, the period of $\mathbf{e}_i$ equals the order of $v(\mathbf{e}_i)$.*

Proof. Denote the period of $\mathbf{e}_i$ by p and the order of $v(\mathbf{e}_i)$ by r. Also, for brevity, let $\mathbf{v} = v(\mathbf{e}_i)$.

First we show that p divides r. By Lemma 6 (with $q = 0$), there is a $k_0 \in \mathbb{N}$ for which

$$\mathcal{C}(\boldsymbol{T}^{kr}(\mathbf{v})) = \mathcal{C}(\mathbf{v})$$

for all $k > k_0$. Since $\mathbf{e}_i \in \mathcal{C}(\mathbf{v})$, then whenever $k > k_0$, we have that $\mathbf{e}_i \in \mathcal{C}(\boldsymbol{T}^{kr}(\mathbf{v}))$, meaning that kr is a return time, so p divides kr. If we choose k to be a suitably large prime number, it follows that p divides r.

We claim that r divides p as well. For this, note that if q is a return time,

$$\mathbf{e}_i \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{e}_i)) \subset \mathcal{C}(\boldsymbol{T}^q(\mathcal{C}(\mathbf{v}))).$$

By Lemma 2,

$$\mathcal{C}(\boldsymbol{T}^q(\mathcal{C}(\mathbf{v}))) \subset \mathcal{C}(\mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))) = \mathcal{C}(\boldsymbol{T}^q(\mathbf{v})).$$

So $\mathbf{e}_i \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ for any return time q.

By Corollary 3, the carriers of the vertices of L are disjoint, so $\mathbf{e}_i \in \mathcal{C}(\boldsymbol{T}^q(\mathbf{v}))$ if and only if $\boldsymbol{T}^q(\mathbf{v}) = \mathbf{v}$. By the definition of the order, this happens if and only if r divides q. This tells us that if q is a return time, then r divides q, meaning that r is a common divisor of the return times of $\mathbf{v}$. By the definition of the greatest common divisor (or rather by one of its basic properties), this means that r divides p, the greatest common divisor of the return times.

Since both p and r are positive integers, and since p divides r and r divides p, then $r = p$. □

Another lemma is useful before we describe ergodic classes and subclasses geometrically.

Lemma 7. *Let $\mathbf{e}_i \in \mathcal{V}(\Delta)$ be ergodic with period r, and let $\mathbf{x} \in \Delta$ be a distribution that is within the limit vertex class of $\mathbf{e}_i$. Also let $q \in \mathbb{N}$. Then*

$$\lim_{k\to\infty} \boldsymbol{T}^{kr+q}(\mathbf{x}) = \boldsymbol{T}^q(v(\mathbf{e}_i)).$$

Proof. For brevity, let $\mathbf{v} = v(\mathbf{e}_i)$. By Theorem 6, the order of $\mathbf{v}$ equals r. Since $\mathbf{x}$ is within the limit vertex class of $\mathbf{e}_i$, then $\mathbf{x} \in \mathcal{C}(\mathbf{v})$. By Corollary 2, we know that for any $q \in \mathbb{N}$

$$\bigcap_{k=0}^{\infty} \boldsymbol{T}^{kr+q}(\mathcal{C}(\mathbf{v})) = \boldsymbol{T}^q(\mathbf{v}).$$

By Lemma 5, this means that

$$\lim_{k\to\infty} \boldsymbol{T}^{kr+q}(\mathbf{x}) = \boldsymbol{T}^q(\mathbf{v})$$

for all $\mathbf{x} \in \mathcal{C}(\mathbf{v})$, which proves the lemma. □

Corollary 7. *Let $\mathbf{e}_i, \mathbf{e}_j \in \mathcal{V}(\Delta)$ be ergodic states. Then $\mathbf{e}_i$ and $\mathbf{e}_j$ are in the same ergodic subclass if and only if they are in the same limit vertex class. Also, $\mathbf{e}_i$ and $\mathbf{e}_j$ are in the same ergodic class if and only if they are in the same orbit class.*

Proof. By Corollary 6, since $\mathbf{e}_i$ and $\mathbf{e}_j$ are ergodic states, they are in $\mathcal{C}(L)$. To identify the orbit classes as the ergodic classes, suppose that $\mathbf{e}_i$ and $\mathbf{e}_j$ are in the same ergodic class. This is true if and only if $\mathbf{e}_i \to \mathbf{e}_j$. By Proposition 8, $\mathbf{e}_i \to \mathbf{e}_j$ if and only if $\mathbf{e}_j \in \boldsymbol{T}^q(v(\mathbf{e}_i))$ for some $q \in \mathbb{Z}^+$, which holds if and only if $\mathbf{e}_j \in \bigcup_{q=0}^{r-1} \mathcal{C}(\boldsymbol{T}^q(v(\mathbf{e}_i)))$. This condition is equivalent to $\mathbf{e}_j \in \mathcal{C}(\boldsymbol{T}^k(v(\mathbf{e}_i)))$ for some k, which is the same as the condition that $v(\mathbf{e}_i)$ and $v(\mathbf{e}_j)$ are in the same orbit. This is equivalent to $\mathbf{e}_i$ and $\mathbf{e}_j$ being in the same orbit class.

To see that the limit vertex classes satisfy the defining properties of ergodic subclasses, first note that orbit classes (which we have shown are the same as ergodic classes) are partitioned into limit vertex classes. Also, $\boldsymbol{T}$ acts cyclically on the vertices in a limit vertex class and $\boldsymbol{T}(\mathcal{C}(v(\mathbf{e}_i))) \subset \mathcal{C}(\boldsymbol{T}(v(\mathbf{e}_i)))$ by Lemma 2, so $\boldsymbol{T}$ maps distributions within a given limit vertex class to distributions within the next limit vertex class in the orbit. The remaining defining convergence property for ergodic subclasses follows from Lemma 7. □

Now that we have characterized ergodic classes and subclasses geometrically, we can prove another main result.

Theorem 7. *If $\mathbf{e}_i, \mathbf{e}_j \in \mathcal{V}(\Delta)$ are ergodic states in the same ergodic class, then they have the same period.*

Proof. Since $\mathbf{e}_i$ is ergodic, then $\mathbf{e}_i \in \mathcal{C}(v(\mathbf{e}_i))$. Since $\mathbf{e}_j$ is in the same ergodic class, then it is in the $\mathcal{C}(\boldsymbol{T}^q(v(\mathbf{e}_i)))$ for some $q \in \mathbb{N}$. Since $v(\mathbf{e}_i)$ and $\boldsymbol{T}^q(v(\mathbf{e}_i))$ are in the same orbit under $\boldsymbol{T}$, they have the same order. By Theorem 6, $\mathbf{e}_i$ and $\mathbf{e}_j$ have the same period. □

6. From intersections to sequences

So far, we have given many results in terms of intersections of sets, but of greater interest in the theory of Markov chains are the limits of distributions under repeated iterations of $\boldsymbol{T}$. We now examine the implications of the main results so far for such limits.

By the definitions of ergodic and transient states, the vertices of Δ are divided into ergodic and transient states. This means that any distribution $\mathbf{x} \in \Delta$ can be written uniquely as a (nonnegative) linear combination of ergodic states plus a (nonnegative) linear combination of transient states:

$$\mathbf{x} = \mathbf{x}_{\text{erg}} + \mathbf{x}_{\text{trans}},$$

where $\mathbf{x}_{\text{erg}}$ and $\mathbf{x}_{\text{trans}}$ are the orthogonal projections onto the spans (in $\mathbb{R}^n$) of the ergodic and transient states.

Theorem 8. *Let $\boldsymbol{T} : \mathbb{R}^n \to \mathbb{R}^n$ be a stochastic linear transformation, and let $\mathbf{x} \in \Delta$. Then*

$$\lim_{k\to\infty} |\boldsymbol{T}^k(\mathbf{x})_{\text{trans}}| = 0.$$

Proof. For any $\varepsilon > 0$, let U be the open set defined by

$$U = \{\mathbf{p} \in \mathbb{R}^n \mid \text{Dist}(\mathbf{p}, L) < \varepsilon\}.$$

Since $L \subset U$, then by Lemma 4, there exists a $k_0 \in \mathbb{Z}^+$, such that $\boldsymbol{T}^k(\Delta) \subset U$ for all integers $k > k_0$. And if $\boldsymbol{T}^k(\Delta) \subset U$, then in particular $\boldsymbol{T}^k(\mathbf{x}) \in U$.

However, $\text{Dist}(\mathbf{p}, L)$ is just $|\mathbf{p}_{\text{trans}}|$, since $\mathbf{p}_{\text{trans}}$ is the orthogonal projection onto the subspace spanned by the transient states, which is the orthogonal complement of the subspace spanned by the ergodic states. So if $\boldsymbol{T}^k(\mathbf{x}) \subset U$ for all $k > k_0$, then $|\boldsymbol{T}^k(\mathbf{x})_{\text{trans}}| < \varepsilon$ for all $k > k_0$, which proves the theorem. □

This leads us to another of the main results for the paper.

Theorem 9. *If $\mathbf{e}_i \in \mathcal{V}(\Delta)$ is transient and $\mathbf{x} \in \Delta$, then*

$$\lim_{k\to\infty} \boldsymbol{T}^k(\mathbf{x}) \cdot \mathbf{e}_i = 0.$$

In traditional language, this states that

$$\lim_{k\to\infty} \text{Prob}(\mathbf{e}_i \text{ in Stage k} \mid \text{initial distribution } \mathbf{x}) = 0.$$

Proof. If we write $\boldsymbol{T}^k(\mathbf{x}_{\text{trans}})$ as

$$\boldsymbol{T}^k(\mathbf{x}_{\text{trans}}) = \alpha_i \mathbf{e}_i + \sum_{\text{other transient } \mathbf{e}_j} \alpha_j \mathbf{e}_j,$$

then by the generalized Pythagorean theorem and using the nonnegativity of the terms involved,

$$|\boldsymbol{T}^k(\mathbf{x}_{\text{trans}})|^2 = \alpha_i^2 + \left| \sum_{\text{other transient } \mathbf{e}_j} \alpha_j \mathbf{e}_j \right|^2 \geq \alpha_i^2 = (\boldsymbol{T}^k(\mathbf{x}) \cdot \mathbf{e}_i)^2,$$

so $0 \leq \boldsymbol{T}^k(\mathbf{x}) \cdot \mathbf{e}_i \leq |\boldsymbol{T}^k(\mathbf{x}_{\text{trans}})|$. By Theorem 8, the largest term goes to zero as $k \to \infty$, so by the sandwich theorem for limits, so does the middle term, which proves the theorem. □

The following theorem tells us exactly which point in L an initial distribution within ergodic states will converge to, proceeding by suitable-sized jumps.

Theorem 10. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be the vertices of L. Suppose that the set of periods for the states is $\{r_1, \ldots, r_m\}$, and let r be the least common multiple of $r_1, \ldots, r_m$. Let $\mathbf{x} \in \Delta$ be an initial distribution that is within ergodic states, meaning that we can expand $\mathbf{x}$ uniquely as a convex combination of ergodic states:*

$$\mathbf{x} = \sum_{\mathbf{e}_j \in \mathcal{C}(L)} \alpha_j \mathbf{e}_j,$$

where $\alpha_1, \ldots, \alpha_n$ are nonnegative real numbers that sum to 1. Then for any $q \in \mathbb{N}$:

$$\lim_{k \to \infty} \boldsymbol{T}^{kr+q}(\mathbf{x}) = \sum_{\mathbf{e}_j \in \mathcal{C}(L)} \alpha_j \boldsymbol{T}^q(v(\mathbf{e}_j)).$$

Proof. Simply use that $\boldsymbol{T}$ is a linear transformation and apply Theorem 7 to each term. □

Other than the fact that they die out in the long run (or equivalently get arbitrarily close to L), not much can be said about the convergence of initial distributions that involve transient states. As many basic examples show, such initial distributions don't necessarily converge at all.

There are two particularly useful types of Markov chains that are guaranteed to converge to a unique stationary distribution. We prove this now for each type from a geometric point of view.

Theorem 11. *If a Markov chain is irreducible and aperiodic, then there exists a unique distribution $\mathbf{v} \in \Delta$ such that $\boldsymbol{T}(\mathbf{v}) = \mathbf{v}$. This distribution also has the property that*

$$\lim_{k \to \infty} \boldsymbol{T}^k(\mathbf{x}) = \mathbf{v}$$

for all $\mathbf{x} \in \Delta$.

Proof. Since the Markov chain is irreducible, then there can be no transient states because ergodic states don't communicate with transient states. So all states are ergodic, which means that $\mathcal{C}(L) = \Delta$.

Also, since the Markov chain is irreducible, all states must comminuicate. By Proposition 8, this means that all states belong to the same orbit class, so $v(\mathbf{e}_i)$ are in the same orbit for all i. But since the Markov chain is aperiodic, the order of this orbit must be 1 by Theorem 6. Putting these together, we see that the vertices of L are partitioned into a single orbit with exactly 1 element, so L has only one vertex, so L is a single point $\mathbf{v}$.

Since $\boldsymbol{T}(L) = L$, it follows that $\boldsymbol{T}(\mathbf{v}) = \mathbf{v}$. Also, any other fixed point of $\boldsymbol{T}$ would necessarily be contained in L (by the definition of L and by other results), so $\boldsymbol{T}$ has a unique fixed point in Δ.

Also, since $v(\mathbf{e}_j) = \mathbf{v}$ for all j and $\boldsymbol{T}^q(\mathbf{v}) = \mathbf{v}$ for all $q \in \mathbb{N}$, applying Theorem 10 completes the proof of the theorem. □

Theorem 12. *If the k-th power of the transition matrix of a Markov chain has all nonzero entries for some positive integer k, then exists a unique distribution $\mathbf{v} \in \Delta$*

such that $\boldsymbol{T}(\mathbf{v}) = \mathbf{v}$. This distribution also has the property that

$$\lim_{k\to\infty} \boldsymbol{T}^k(\mathbf{x}) = \mathbf{v}$$

for all $\mathbf{x} \in \Delta$.

Proof. If the k-th power of the transition matrix of a Markov chain has all nonzero entries for some positive integer k, geometrically this tells us that $\boldsymbol{T}^k(\Delta) \subset \text{Interior}(\Delta)$. This means that any vertex $\mathbf{v} \in L \subset \boldsymbol{T}^k(\Delta)$ is in Interior(Δ), so $\mathcal{C}(\mathbf{v}) = \Delta$. Since the carriers of the vertices of L are disjoint by Proposition 3, then L has only a single vertex and so is a single point. Applying Theorem 10 as in the proof of Theorem 11 completes the proof. □

7. Conclusion

In this paper we have shown that by viewing them in geometric terms, the main convergence theorems for Markov chains can not only be stated and described more easily, but they can also be proved without invoking any major theorems and without using techniques beyond basic linear algebra, point set topology, and a little bit of abstract algebra. Also, although we have not done so in this paper, pictures of the ways that Markov chains converge can also be drawn. Of course, this is only literally true for Markov chains with three or fewer states, but many of the most salient features can be seen by using these as schematic diagrams for the behavior of Markov chains with more states.

Bibliography

[1] N. Pullman (1965) Geometry of Markov Chains. *Canad. Math. Bull.* **8**, 345-358.

University of Puget Sound, Department of Mathematics, Tacoma, WA 98416-1031
E-mail address: jbernhard@pugetsound.edu