

Characterisation of ergodic upper transition operators

Filip Hermans, Gert de Cooman

SYSTeMS Research Group, Ghent University, Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium

Abstract

We study ergodicity for upper transition operators: bounded, sub-additive and non-negatively homogeneous transformations of finite-dimensional linear spaces. Ergodicity provides a necessary and sufficient condition for Perron-Frobenius-like convergence behaviour for upper transition operators. It can also be characterised alternatively: (i) using a coefficient of ergodicity, and (ii) using accessibility relations. The latter characterisation states that ergodicity is equivalent with there being a single maximal communication (or top) class that is moreover regular and absorbing. We present an algorithm for checking these conditions that is linear in the dimension of the state space for the number of evaluations of the upper transition operator.

Keywords: upper transition operators, imprecise Markov chain, ergodicity, Perron-Frobenius

1. Introduction

Throughout the paper, \mathcal{X} denotes a finite non-empty set of elements that we also refer to as *states*, and $\mathcal{L}(\mathcal{X})$ is the set of all real-valued maps on \mathcal{X} . We provide the finite-dimensional linear space $\mathcal{L}(\mathcal{X})$ with the supremum norm $\|\cdot\|_\infty$, or with the topology of uniform convergence, so the result is a Banach space. Uniform and pointwise convergence coincide on this finite-dimensional space. Given f and g in $\mathcal{L}(\mathcal{X})$, we write $f \geq g$ if $f(x) \geq g(x)$ for all $x \in \mathcal{X}$. We also define $\min f := \min \{f(x) : x \in \mathcal{X}\}$ and $\max f := \max \{f(x) : x \in \mathcal{X}\}$.

Definition 1. An *upper transition operator* on $\mathcal{L}(\mathcal{X})$ is a transformation $T: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X})$ that has the following properties:

- | | |
|----------------------------------|---|
| T1. $\min f \leq Tf \leq \max f$ | <i>T is bounded;</i> |
| T2. $T(f + g) \leq Tf + Tg$ | <i>T is sub-additive;</i> |
| T3. $T(\lambda f) = \lambda Tg$ | <i>T is non-negatively homogeneous;</i> |

for arbitrary f, g in $\mathcal{L}(\mathcal{X})$ and real $\lambda \geq 0$.

Email addresses: filip.hermans@UGent.be (Filip Hermans), gert.decooman@UGent.be (Gert de Cooman)

Any upper transition operator T automatically also satisfies the following interesting properties:

- T4. $T(f + \mu) = Tf + \mu$ T is *constant-additive*;
- T5. if $f \leq g$ then $Tf \leq Tg$ T is *order-preserving*;
- T6. if $f_n \rightarrow f$ then $Tf_n \rightarrow Tf$ T is *continuous*;
- T7. $Tf + T(-f) \geq 0$ T is *upper-lower consistent*;

for arbitrary f, g, f_n in $\mathcal{L}(\mathcal{X})$ and real μ . Clearly, for any n in the set of natural numbers (with zero) \mathbb{N}_0 , T^n is an upper transition operator as well.

Properties T4 and T5 define a topical map [6, Sec. 4]. It is easy to see [6, Prop. 4.1] that any topical map is also non-expansive under the supremum norm: for all f and g in $\mathcal{L}(\mathcal{X})$,

- T8. $\|Tf - Tg\|_\infty \leq \|f - g\|_\infty$ T is *non-expansive*.

A very useful result for non-expansive maps by Sine [11, Theo. 1] and Nussbaum [8, 9]¹ states that for every element f of the finite-dimensional domain of a non-expansive transformation T , there is some natural number p such that the sequence $T^{np}f$ converges. More importantly, Sine proves that we can find a finite ‘period’ p *common* to all maps f in the domain $\mathcal{L}(\mathcal{X})$. This means that, for any f , the set $\omega_T(f)$ of limit points of the set of iterates $\{T^n f : n \in \mathbb{N}\}$ has a number of elements $|\omega_T(f)|$ that divides this p .² T is cyclic on $\omega_T(f)$, with period $|\omega_T(f)|$ (and therefore also with period p). Lemmens and Scheutzw [6, Theo. 5.2] managed to prove that an upper bound for the common periods of all topical functions $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is $\binom{n}{\lfloor n/2 \rfloor}$. This upper bound is tight in the sense that there is always at least one topical function that has this bound as its smallest common period. However, Akian and Gaubert [1, Cor. 5.6] have shown that for convex maps that are monotone and non-expansive, this bound is equal to the maximal order of the permutation group. This is given by Landau’s function g for which $\ln g(n) \sim c_1 \sqrt{n \ln n}$, whereas $\ln \binom{n}{\lfloor n/2 \rfloor} \sim c_2 n$, for some constants $c_1, c_2 > 0$, as $n \rightarrow \infty$.

In Sec. 3 we use these ideas to introduce ergodicity for upper transition operators, and to explain its link with so-called Perron-Frobenius conditions. That there is such a link has already been established by Akian and Gaubert [1, Theo. 1.1] for a more general class of operators. The goal of Akian and Gaubert was to determine combinatorial bounds for the orbit lengths of order preserving, convex and sup-norm non-expansive maps. These upper bounds involve the notion of a critical graph. It is shown in [1, Theo. 6.6], that these bounds are tight when the map is piecewise affine. Moreover, in this case, Akian and Gaubert give an algorithm to compute the critical graph. In this paper, we assume in addition to the general assumptions made by Akian and Gaubert, that the map is non-negatively homogeneous and we address the case where all periodic orbits have length one. For this class of maps, we show that the piecewise affine condition can

¹Nussbaum found and closed a gap in Sine’s argument.

² $|A|$ denotes the cardinality of a set A and \mathbb{N} is the set of natural numbers (without zero).

be dispensed with for the critical graph bound to be tight. In Sec. 7 we compare our approach to the critical graph method used by Akian and Gaubert.

In addition, using the alternative characterisation of ergodicity developed in Sec. 3, we are able in Sec. 5 to avoid the critical—in terms of computational complexity—step of Akian and Gaubert’s algorithm in [1, Sec. 6.3]: the computation of the subdifferential, which relies heavily on extreme points. Our newly designed algorithm is linear in the dimension of the state space, where the evaluation of the transition map is considered as an oracle.

In Sec. 6 we prove that ergodicity is equivalent to a contraction property in Hilbert’s seminorm which is the approach that was previously followed by Škulj and Hable [12]. We explain the advantages and disadvantages of characterisation of ergodicity in terms of a coefficient of ergodicity.

2. Upper transition operators and imprecise Markov chains

Upper transition operators are introduced by De Cooman [2, Sec. 3] when describing imprecise Markov chains. These imprecise Markov chains are random processes where prior and transition beliefs are described in terms of Walley’s [13, Sec. 2.3.3] coherent upper previsions. In this framework, $T^k \mathbb{I}_A(x)$ can be interpreted as the upper probability and $1 - T^k \mathbb{I}_{A^c}(x)$ as the lower probability to go from state x in k steps to some state in A .

To see where this interpretation comes from, consider $\bar{P}(f|x) := Tf(x)$ and observe that $\bar{P}(\cdot|x)$ is a bounded, sub-additive, non-negatively homogeneous, real-valued functional. This type of functional is exactly what Walley [13] calls a *coherent upper (conditional) prevision*. Because $\bar{P}(\cdot|x)$ is order-preserving (T5), constant additive (T4), convex (T2+T3) and non-negatively homogeneous (T3), it follows from Legendre-Fenchel duality, that $\bar{P}(f|x)$ can be written as

$$\bar{P}(f|x) = \max \{p \cdot f : p \in \mathcal{P}_x\},$$

where \mathcal{P}_x is a compact convex set of stochastic vectors, also known as a *credal set*. The upper transition operator Tf can now be seen as the Cartesian product of the upper previsions over all states. If given a prior upper prevision \bar{P}_0 corresponding to a credal set \mathcal{P}_0 :

$$\bar{P}_0(f) = \max \{p_0 \cdot f : p_0 \in \mathcal{P}_0\},$$

then it follows almost immediately that

$$\bar{P}_0(Tf) = \max \{p_0 \cdot M \cdot f : p_0 \in \mathcal{P}_0 \text{ and } M \in \mathcal{T}\},$$

where

$$\mathcal{T} := \left\{ M \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}} : (\forall x \in \mathcal{X})(M_{x,\cdot} \in \mathcal{P}_x) \right\}. \quad (1)$$

Here, M is a stochastic matrix where the x -th row, $M_{x,\cdot}$, is a probability distribution over the states at a time $k+1$, conditional on the chain being in state x at time k . Therefore, we can interpret M as a transition matrix of a finite-state and discrete-time Markov chain. When considering iterations of the map, then we see that

$$\bar{P}_0(T^k f) = \max \left\{ p_0 \cdot M^{(1)} \cdot \dots \cdot M^{(k)} \cdot f : p_0 \in \mathcal{P}_0 \text{ and } M^{(j)} \in \mathcal{T} \right\}.$$

Generally speaking, therefore, an upper transition operator effects robust inference for a set of not necessarily stationary Markov chains. For more details, we refer to [2, 4, 12].

3. Perron-Frobenius condition for upper transition operators

In this section we introduce the notion of ergodicity for upper transition operators and lay bare the link with the Perron-Frobenius theorem. We allow ourselves to be inspired by corresponding notions for non-stationary Markov chains [10, p. 136] and Markov set chains [4] to lead us to the following definition of ergodicity.

Definition 2 (Ergodicity). An upper transition operator T on $\mathcal{L}(\mathcal{X})$ is called *ergodic* if for all $f \in \mathcal{L}(\mathcal{X})$, $\lim_{n \rightarrow \infty} T^n f$ exists and is a constant function.

This definition of ergodicity is not exactly the one more commonly encountered in probability or dynamical systems theory, where ergodicity usually refers to the special properties of an invariant measure. Here, ergodicity corresponds to what is usually called “ergodic + aperiodic” in the Markov chain setting.

Consider any $f \in \mathcal{L}(\mathcal{X})$. Ergodicity of an upper transition operator T not only means that the sequence $T^n f$ converges, so $\omega_T(f)$ is a singleton $\{\xi_f\}$, but also that this limit ξ_f is a constant function. Observe that by T6, ξ_f is a fixed point for all T^k : $T^k \xi_f = \xi_f$ and therefore $\xi_{T^k f} = \xi_f$ for all $k \in \mathbb{N}$. If we denote the constant value of ξ_f by $\bar{E}_T(f)$, then this defines a real functional \bar{E}_T on $\mathcal{L}(\mathcal{X})$. This functional is an *upper expectation*: it is bounded, sub-additive and non-negatively homogeneous [compare with T1–T3]. It is T -invariant in the sense that $\bar{E}_T \circ T = \bar{E}_T$, and it is the only such upper expectation. This shows that our definition of ergodicity is nevertheless in line with the concept used in systems theory.

Definition 3. An upper transition operator T on $\mathcal{L}(\mathcal{X})$ is called *Perron–Frobenius-like* if there is some real functional \bar{E}_∞ on $\mathcal{L}(\mathcal{X})$ such that $\lim_{n \rightarrow \infty} \bar{E}(T^n f) = \bar{E}_\infty(f)$ for all upper expectations \bar{E} on $\mathcal{L}(\mathcal{X})$ and all $f \in \mathcal{L}(\mathcal{X})$, or in other words, if the sequence of upper expectations $\bar{E} \circ T^n$ converges to some limit that does not depend on the initial value \bar{E} .

As an immediate result, conditions for ergodicity of upper transition operators are conditions for a Perron–Frobenius-like theorem for such transformations to hold.

Theorem 1 (Perron–Frobenius). An upper transition operator T is *Perron–Frobenius-like* if and only if it is *ergodic*, and in that case $\bar{E}_\infty = \bar{E}_T$.

PROOF. Sufficiency. Suppose T is ergodic. Then using the notations established above, $T^n f \rightarrow \xi_f$ and therefore $\bar{E}(T^n f) \rightarrow \bar{E}(\xi_f)$ because any upper expectation \bar{E} is continuous [compare with T6]. Observe that, since any upper expectation \bar{E} is constant-additive [compare with T4 and T1], $\bar{E}(\xi_f) = \bar{E}_T(f)$. Hence $\bar{E} \circ T^n \rightarrow \bar{E}_T$, and therefore T is Perron–Frobenius-like, with $\bar{E}_\infty = \bar{E}_T$.

Necessity. Suppose that T is Perron–Frobenius-like, with limit upper expectation \bar{E}_∞ . Fix any $x \in \mathcal{X}$, and consider the upper expectation \bar{E}_x defined by $\bar{E}_x(f) := f(x)$ for all $f \in \mathcal{L}(\mathcal{X})$. Then by assumption $T^n f(x) = \bar{E}_x(T^n f) \rightarrow \bar{E}_\infty(f)$. Since this holds for all $x \in \mathcal{X}$, we see that T is ergodic with $\bar{E}_T = \bar{E}_\infty$.

It follows from the discussion in Sec. 1 that $\bigcup_{f \in \mathcal{L}(\mathcal{X})} \omega_T(f)$ is the set of all periodic points of T —a *periodic point* being an element $f \in \mathcal{L}(\mathcal{X})$ for which there is some $n \in \mathbb{N}$ for which $T^n f = f$. Because of T4, this set contains all constant maps. We now see that for T to be ergodic, this set cannot contain any other maps.

Proposition 2. *An upper transition operator T is ergodic if and only if all of its periodic points are constant maps.*

4. Characterisation of ergodicity

We now turn to the issue of determining in practise whether an upper transition operator is ergodic. In the case of finite-state, discrete-time Markov chains, a nice approach to deciding upon ergodicity was given by Kemeny and Snell [5, Sec. 1.4]. It is based on the notion of an accessibility relation. This is a binary (weak order) relation on a set of states \mathcal{X} that captures whether it is possible to go from one state to another in a finite number of steps. We now intend to show that it is possible to associate an accessibility relation with an upper transition operator, and that this relation provides us with an intuitive interpretation of the notion of ergodicity in terms of accessibility. We refer to [2] for a detailed discussion of accessibility relations and their connections with upper transition operators.

Definition 4. Consider an upper transition operator T on $\mathcal{L}(\mathcal{X})$, and two states x and y in \mathcal{X} . We say that y is *accessible*³ from x in n steps (notation: $x \xrightarrow{n} y$) if $T^n \mathbb{I}_{\{y\}}(x) > 0$. We say that state y is *accessible* from state x (notation: $x \rightarrow y$) if $T^n \mathbb{I}_{\{y\}}(x) > 0$ for some $n \in \mathbb{N}_0$. We say that x and y *communicate* (notation: $x \leftrightarrow y$) if both $x \rightarrow y$ and $y \rightarrow x$.

The relation \rightarrow is a weak order (reflexive and transitive), and consequently \leftrightarrow is an equivalence relation. The equivalence classes for this relation are called *communication classes*: maximal subsets of \mathcal{X} for which every element has access to any other element. The accessibility relation induces a partial order on these communication classes.

In the case of finite-state, discrete-time Markov chains, this partial order gives us clues about the ergodicity of the Markov chain. For such a Markov chain to be ergodic, it is necessary and sufficient that [2] it is *top class regular*, meaning that: (i) there is only one *maximal* or *undominated* communication class—elements of a maximal communication class have no access to states not in that class—, in which case we call this unique maximal class \mathcal{R} the *top class*; and (ii) the top class \mathcal{R} should be regular, meaning that after some time k , all elements of this class become accessible to each other in any number of steps: for all x and y in \mathcal{R} and for all $n \geq k$, $x \xrightarrow{n} y$.

For upper transition operators, it turns out that top class regularity is a necessary condition for ergodicity. However, top class regularity is by itself not a sufficient condition: we need some guarantee that the top class will eventually be reached—a requirement that is automatically fulfilled in finite-state discrete-time Markov chains.

³If the upper transition operator is interpreted in terms of upper previsions (see also [2]), then the term accessible might be a tad presumptuous. A more accurate term would be *possibly accessible* or *not excluded from being accessible*.

Proposition 3. *An upper transition operator T is ergodic if and only if it is regularly absorbing, meaning that it satisfies the following properties:*

(TCR) *it is top class regular:*

$$\mathcal{R} := \left\{ x \in \mathcal{X} : (\exists n \in \mathbb{N})(\forall k \geq n) \min T^k \mathbb{I}_{\{x\}} > 0 \right\} \neq \emptyset,$$

(TCA) *it is top class absorbing: with $\mathcal{R}^c := \mathcal{X} \setminus \mathcal{R}$,*

$$(\forall y \in \mathcal{R}^c)(\exists n \in \mathbb{N}) T^n \mathbb{I}_{\mathcal{R}^c}(y) < 1.$$

For a proof that (TCR) is equivalent to $\mathcal{R} \neq \emptyset$, we refer to [2, Prop. 4.3]. (TCA) means that for every element y not in the top class, there is some finite number of steps n after which the top class can be reached with a strictly positive *lower* probability $1 - T^n \mathbb{I}_{\mathcal{R}^c}(y)$.

PROOF. (TCR) \wedge (TCA) \Rightarrow (ER). Consider any fixed point ξ of T^k , where $k \in \mathbb{N}$ and observe, by T5 and T4, that $\min \xi \leq \min T \xi \leq \min T^2 \xi \leq \dots \leq \min T^k \xi = \min \xi$ whence for any $p \in \mathbb{N}$,

$$\min \xi = \min T^p \xi \quad \text{and similarly} \quad \max \xi = \max T^p \xi. \quad (2)$$

We infer from Prop. 2 that we have to show that ξ is constant. Using T5, T4, T3 and Eq. 2 we construct from $T^p \xi \geq \min T^p \xi + [T^p \xi(x) - \min T^p \xi] \mathbb{I}_{\{x\}} = \min \xi + [T^p \xi(x) - \min \xi] \mathbb{I}_{\{x\}}$ the following inequality, which holds for all $n, p \in \mathbb{N}_0$ and all $x \in \mathcal{X}$:

$$T^n \xi \geq \min \xi + [T^p \xi(x) - \min \xi] T^n \mathbb{I}_{\{x\}}.$$

Hence, by taking the minimum on both sides of this inequality and using Eq.(2), we find that

$$0 \geq [T^p \xi(x) - \min \xi] \min T^n \mathbb{I}_{\{x\}}.$$

We infer from (TCR) that by taking n large enough, we can ensure that $\min T^n \mathbb{I}_{\{x\}} > 0$ whence for any $p \in \mathbb{N}_0$ and $x \in \mathcal{R}$

$$0 = [T^p \xi(x) - \min \xi],$$

so we already find that $T^p \xi(x) = \min \xi$ for all $p \in \mathbb{N}_0$ and $x \in \mathcal{R}$.

If there is some $p \in \mathbb{N}_0$ such that $T^p \xi$ reaches its maximum on \mathcal{R} , then we infer from Eq. (2) that $\max T^p \xi = \max \xi$ which has to be equal to $\min \xi$ to satisfy the inequality, so ξ is indeed constant. Let us therefore assume that the maximum of $T^p \xi$ is not reached in \mathcal{R} . Using T5, T4, T3 and Eq. (2), we construct from $\xi \leq \max \xi - [\max \xi - \max_{x \in \mathcal{R}} \xi(x)] \mathbb{I}_{\mathcal{R}}$ and $-\mathbb{I}_{\mathcal{R}} = \mathbb{I}_{\mathcal{R}^c} - 1$ the following inequality, which holds for all $n \in \mathbb{N}$:

$$T^n \xi \leq \max \xi + \left[\max \xi - \max_{x \in \mathcal{R}} \xi(x) \right] (T^n \mathbb{I}_{\mathcal{R}^c} - 1).$$

By taking the maximum over \mathcal{R}^c on both sides of this inequality and under the made assumption that the maximum is never reached on \mathcal{R} , we get

$$0 = \max_{y \in \mathcal{R}^c} \mathbf{T}^n \xi(y) - \max \xi \leq \left[\max \xi - \max_{x \in \mathcal{R}} \xi(x) \right] \left(\max_{y \in \mathcal{R}^c} \mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}(y) - 1 \right).$$

For each $y \in \mathcal{R}^c$, consider some $n_y \in \mathbb{N}$ such that $\mathbf{T}^{n_y} \mathbb{I}_{\mathcal{R}^c}(y) < 1$, and let $n := \max_{y \in \mathcal{R}^c} n_y$. Then we see that for every $y \in \mathcal{R}^c$:

$$\mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}(y) = \mathbf{T}^{n_y} [(\mathbb{I}_{\mathcal{R}} + \mathbb{I}_{\mathcal{R}^c}) \mathbf{T}^{n-n_y} \mathbb{I}_{\mathcal{R}^c}](y) = \mathbf{T}^{n_y} [\mathbb{I}_{\mathcal{R}^c} \mathbf{T}^{n-n_y} \mathbb{I}_{\mathcal{R}^c}](y) \leq \mathbf{T}^{n_y} \mathbb{I}_{\mathcal{R}^c}(y) < 1.$$

The second equality follows from the fact that $\mathbb{I}_{\mathcal{R}} \mathbf{T}^{n-n_y} \mathbb{I}_{\mathcal{R}^c} = 0$: an element in the top class \mathcal{R} has no access to any element outside of it; and the first inequality follows from $\mathbb{I}_{\mathcal{R}^c} \leq 1$ and T5. But this means that $\max_{y \in \mathcal{R}^c} \mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}(y) - 1 < 0$ and consequently $\max \xi = \max_{x \in \mathcal{R}} \xi(x) = \min \xi$.

(ER) \Rightarrow (TCR) \wedge (TCA). We will use contraposition and show first that $\neg(\text{TCR}) \Rightarrow \neg(\text{ER})$. Then we will show that $\neg(\text{TCA}) \wedge (\text{TCR}) \Rightarrow \neg(\text{ER})$.

$\neg(\text{TCR}) \Rightarrow \neg(\text{ER})$. Not being top class regular means that $\mathcal{R} = \emptyset$, which is equivalent to

$$(\forall x \in \mathcal{X})(\forall n \in \mathbb{N})(\exists k \geq n)(\exists z \in \mathcal{X}) \mathbf{T}^k \mathbb{I}_{\{x\}}(z) = 0.$$

Since we infer from $\mathbb{I}_{\{x\}} \geq 0$ and T1 that $\mathbf{T}^k \mathbb{I}_{\{x\}} \geq 0$, this leads us to conclude that $\liminf_{n \rightarrow \infty} \min \mathbf{T}^n \mathbb{I}_{\{x\}} = 0$. But for any $n \in \mathbb{N}$, $\mathbf{T}^{n+1} \mathbb{I}_{\{x\}} = \mathbf{T}(\mathbf{T}^n \mathbb{I}_{\{x\}}) \geq \min \mathbf{T}^n \mathbb{I}_{\{x\}}$ by T1, and therefore also $\min \mathbf{T}^{n+1} \mathbb{I}_{\{x\}} \geq \min \mathbf{T}^n \mathbb{I}_{\{x\}}$. This implies that the sequence $\min \mathbf{T}^n \mathbb{I}_{\{x\}}$ is non-decreasing, and bounded above [by 1], and therefore convergent. This shows that

$$(\forall x \in \mathcal{X}) \lim_{n \rightarrow \infty} \min \mathbf{T}^n \mathbb{I}_{\{x\}} = 0. \quad (3)$$

We also infer from T1 and T2 that $1 = \mathbf{T}^k \mathbb{I}_{\mathcal{X}} \leq \sum_{x \in \mathcal{X}} \mathbf{T}^k \mathbb{I}_{\{x\}}$. Since the cardinality $|\mathcal{X}|$ of the state space is finite, this means that for all $z \in \mathcal{X}$ and all $n \in \mathbb{N}$ there is some $x \in \mathcal{X}$ such that $\mathbf{T}^n \mathbb{I}_{\{x\}}(z) \geq 1/|\mathcal{X}|$. This tells us that $\max \mathbf{T}^n \mathbb{I}_{\{x\}} \geq 1/|\mathcal{X}|$. Since we can infer from a similar argument as before that the sequence $\max \mathbf{T}^n \mathbb{I}_{\{x\}}$ converges, this tells us that

$$(\forall x \in \mathcal{X}) \lim_{n \rightarrow \infty} \max \mathbf{T}^n \mathbb{I}_{\{x\}} \geq \frac{1}{|\mathcal{X}|}. \quad (4)$$

Combining Eqs. (3) and (4) tells us that $\lim_{n \rightarrow \infty} (\max \mathbf{T}^n \mathbb{I}_{\{x\}} - \min \mathbf{T}^n \mathbb{I}_{\{x\}}) > 0$, so T cannot be ergodic.

$\neg(\text{TCA}) \wedge (\text{TCR}) \Rightarrow \neg(\text{ER})$. Since T is not top class absorbing, we know that there is some $y \in \mathcal{R}^c$ such that $\mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}(y) = 1$ for all $n \in \mathbb{N}$. As the top class \mathcal{R} is non-empty, we know that there is some $x \in \mathcal{R}$, and this x has no access to any state outside the maximal communication class \mathcal{R} : $\mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}(x) = 0$ for all $n \in \mathbb{N}$. Consequently

$$\lim_{n \rightarrow \infty} (\max \mathbf{T}^n \mathbb{I}_{\mathcal{R}^c} - \min \mathbf{T}^n \mathbb{I}_{\mathcal{R}^c}) = 1 - 0 > 0,$$

so T cannot be ergodic.

5. Ergodicity checking in practise

5.1. Checking for top class regularity

Checking for top class regularity directly using the definition would involve calculating for every state x the maps $\mathbb{T}\mathbb{I}_{\{x\}}$, $\mathbb{T}^2\mathbb{I}_{\{x\}}$, \dots , $\mathbb{T}^n\mathbb{I}_{\{x\}}$ until a first number $n = n_x$ is found for which $\min \mathbb{T}^n\mathbb{I}_{\{x\}} > 0$. Unfortunately, it is not clear whether this procedure is guaranteed to terminate after a certain number of iterations, or whether we can stop checking after a fixed number of iterations. In this section, we want to take a closer look at this problem.

The next proposition shows that all the information we need in order to check top class regularity is incorporated in a single application of \mathbb{T} to the atoms of \mathcal{X} .

Proposition 4. *Let \mathbb{T} be an upper transition operator on $\mathcal{L}(\mathcal{X})$, $n \in \mathbb{N}$ and $x, y \in \mathcal{X}$. Then $\mathbb{T}^n\mathbb{I}_{\{y\}}(x) > 0$ if and only if there is some sequence $x_0, x_1, x_2, \dots, x_{n-1}, x_n$ in \mathcal{X} with $x_0 = x$ and $x_n = y$ such that $\mathbb{T}\mathbb{I}_{\{x_{k+1}\}}(x_k) > 0$ for all $k \in \{0, 1, \dots, n-1\}$.*

PROOF. Sufficiency. Fix k, ℓ in \mathbb{N} , and u, v in \mathcal{X} . Since $\mathbb{T}^\ell\mathbb{I}_{\{y\}} = \sum_{z \in \mathcal{X}} \mathbb{I}_{\{z\}} \mathbb{T}^\ell\mathbb{I}_{\{y\}}(z) \geq \mathbb{I}_{\{v\}} \mathbb{T}^\ell\mathbb{I}_{\{y\}}(v)$, it follows from T5 and T3 that $\mathbb{T}^{k+\ell}\mathbb{I}_{\{y\}} \geq \mathbb{T}^k\mathbb{I}_{\{v\}} \mathbb{T}^\ell\mathbb{I}_{\{y\}}(v)$ and therefore $\mathbb{T}^{k+\ell}\mathbb{I}_{\{y\}}(x) \geq \mathbb{T}^k\mathbb{I}_{\{v\}}(x) \mathbb{T}^\ell\mathbb{I}_{\{y\}}(v)$. Applying this inequality repeatedly, we get:

$$\mathbb{T}^n\mathbb{I}_{\{y\}}(x) \geq \prod_{k=0}^{n-1} \mathbb{T}\mathbb{I}_{\{x_{k+1}\}}(x_k)$$

for any sequence $x_0, x_1, x_2, \dots, x_{n-1}, x_n$ in \mathcal{X} with $x_0 = x$ and $x_n = y$. It follows that the left-hand side is positive as soon as all factors on the right-hand side are.

Necessity. We infer using T2 and T3 that

$$\mathbb{T}^n\mathbb{I}_{\{y\}}(x) = \mathbb{T}\left(\sum_{x_1 \in \mathcal{X}} \mathbb{I}_{\{x_1\}} \mathbb{T}^{n-1}\mathbb{I}_{\{y\}}(x_1)\right)(x) \leq \sum_{x_1 \in \mathcal{X}} \mathbb{T}^{n-1}\mathbb{I}_{\{y\}}(x_1) \mathbb{T}\mathbb{I}_{\{x_1\}}(x),$$

and repeating the same argument recursively leads to

$$\mathbb{T}^n\mathbb{I}_{\{y\}}(x) \leq \sum_{\substack{x_0, x_1, \dots, x_{n-1}, x_n \in \mathcal{X} \\ x_0 = x, x_n = y}} \prod_{k=0}^{n-1} \mathbb{T}\mathbb{I}_{\{x_{k+1}\}}(x_k).$$

Since all the factors (and therefore all the terms) on the right-hand side are non-negative by T1 and T5, the positivity of the left-hand side implies that there must be at least one positive term on the right-hand side, all of whose factors must therefore be positive.

This proposition not only implies that the set $\{\mathbb{T}\mathbb{I}_{\{x\}} : x \in \mathcal{X}\}$ completely determines the accessibility relation \rightarrow , but also that it determines the ‘accessibility in n steps’ relation \xrightarrow{n} . In other words, not only the communication and maximal classes can be determined from $\{\mathbb{T}\mathbb{I}_{\{x\}} : x \in \mathcal{X}\}$, but also their regularity.

Let us first recall some notions of graph theory before continuing; see for instance [1] for more details. For two vertices x and y of a graph \mathcal{G} we say that x has access to y if there exists a path from x to y or if $x = y$, and we denote this by $x \xrightarrow{*} y$. A *strongly*

connected component C of the graph $\mathcal{G}|_C$ is the restriction of the \mathcal{G} to the set of nodes C , where C is an equivalence class of the equivalence relation ' $x \xrightarrow{*} y$ and $y \xrightarrow{*} x$ '. Any such equivalence class C (resulting in a strongly connected component) is said to be *final* if no vertex in C has access to any vertex outside C . Finally, the *period* of a strongly connected component C is defined as the greatest common divisor of the lengths of the circuits of $\mathcal{G}|_C$ where a *circuit* is a closed path where no node has been visited more than once.

Definition 5. The *upper accessibility graph* $\overline{\mathcal{G}}(\mathbb{T})$ corresponding to an upper transition operator \mathbb{T} is the directed graph with set of edges $\{(x, y) \in \mathcal{X}^2 : \mathbb{T}\mathbb{I}_y(x) > 0\}$ and set of vertices \mathcal{X} .

The upper accessibility graph we define here, is a special case of the *syntactic digraph* considered in Gaubert and Gunawardena [3, Prop. 2]; in that graph, there is an arc from x to y if and only if the coordinate y of \mathbb{T} depends logically on the x -coordinate in the argument.

It is clear from Prop. 4 that the accessibility relation $\xrightarrow{*}$ of the graph $\overline{\mathcal{G}}(\mathbb{T})$ is exactly the accessibility relation \rightarrow belonging to the upper transition operator \mathbb{T} . This means that checking for the existence of a single top class of \mathbb{T} corresponds to asserting whether there is only one final class \mathcal{R} in $\overline{\mathcal{G}}(\mathbb{T})$. Once we have found the top class \mathcal{R} , we focus on the subgraph $\overline{\mathcal{G}}(\mathbb{T})|_{\mathcal{R}}$ which is the upper accessibility graph $\overline{\mathcal{G}}(\mathbb{T})$ restricted to \mathcal{R} . Prop. 4.2 in [2] tells us that checking for regularity of the top class means that we have to check whether the cyclicity of $\overline{\mathcal{G}}(\mathbb{T})$ is equal to 1.

The relation between \mathbb{T} and its graph $\overline{\mathcal{G}}(\mathbb{T})$ is a purely qualitative one: the exact quantitative value of the upper transition probabilities between two states x and y is not important at all. What is important is whether there is a possible transition between two states. This means that appropriately replacing the upper transition operator \mathbb{T} with a classical, linear transition operator, or its associated transition matrix M , will still lead to the same results.

Definition 6. A stochastic matrix $M \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$ represents an upper transition operator \mathbb{T} on $\mathcal{L}(\mathcal{X})$ if $M_{x,y} > 0 \Leftrightarrow \mathbb{T}\mathbb{I}_{\{y\}}(x) > 0$ for all x and y in \mathcal{X} .

It is clear that any stochastic matrix M that represents \mathbb{T} will result into the same graph $\overline{\mathcal{G}}(\mathbb{T})$ and will therefore lead to the same conclusions with respect to top class regularity. For stochastic matrices however, a final class corresponds to a stochastic submatrix and aperiodicity corresponds to the absence of eigenvalues with modulus 1 apart from 1 itself.

Proposition 5 (Top class regularity). Consider an upper transition operator \mathbb{T} . Then the following statements are equivalent: (i) \mathbb{T} is top class regular; (ii) M represents \mathbb{T} and is regular; (iii) M represents \mathbb{T} and M has exactly one eigenvalue with modulus 1; and (iv) $\overline{\mathcal{G}}(\mathbb{T})$ has exactly one final class \mathcal{R} and $\overline{\mathcal{G}}(\mathbb{T})|_{\mathcal{R}}$ has cyclicity 1.

Example 1. Let $\mathcal{X} := \{x, y\}$ and $\mathbb{T}f := f(x)\mathbb{I}_{\{x\}} + \max\{f(x), f(y)\}\mathbb{I}_{\{y\}}$ for all $f \in \mathcal{L}(\mathcal{X})$. Then $\mathbb{T}\mathbb{I}_{\{x\}} = \mathbb{I}_{\mathcal{X}}$ whence $x \xrightarrow{1} x$ and $y \xrightarrow{1} x$ and $\mathbb{T}\mathbb{I}_{\{y\}} = \mathbb{I}_{\{y\}}$ whence $y \xrightarrow{1} y$. The graph $\overline{\mathcal{G}}(\mathbb{T})$ is then given by



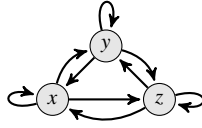
Clearly $\{x\}$ is the unique final strongly connected component of $\overline{\mathcal{G}}(T)$ and as it is a singleton, it has cyclicity one. We conclude that T is top class regular.

In the next example we focus on a simple upper transition operator that is not piecewise affine. It does not therefore fall within the scope of Akian and Gaubert's algorithm.

Example 2. Consider the map

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3: f \rightarrow \bar{f} + \frac{\|f - \bar{f}\|_2}{\sqrt{3}} \begin{pmatrix} \alpha_x \\ \alpha_y \\ \alpha_z \end{pmatrix}$$

where $\bar{f} := (f_x + f_y + f_z)/3$ for $f = (f_x \ f_y \ f_z)^T$ and the parameters α_x , α_y and α_z are any real numbers in $[0, 1/\sqrt{2}]$. It is not difficult to check that this T is indeed an upper transition operator, but it is obviously not piecewise affine. Independently of the value of α_x , α_y and α_z , the upper accessibility graph of this map is given by:



The entire graph is strongly connected, and it has cyclicity one. This implies that T is not only top class regular, but also ergodic, according to Prop. 3.

5.2. Checking for top class absorption

We now present a computationally cheap procedure to check for top class absorption.

Proposition 6 (Top class absorption). *Let T be an upper transition operator with regular top class \mathcal{R} . Consider the nested sequence of subsets of \mathcal{R}^c defined by the iterative scheme:*

$$A_0 := \mathcal{R}^c$$

$$A_{n+1} := \{a \in A_n: T\mathbb{I}_{A_n}(a) = 1\}, \quad n \geq 0.$$

After $k \leq |\mathcal{R}^c|$ iterations, we reach $A_k = A_{k+1}$. Then T is top class absorbing if and only if $A_k = \emptyset$.

PROOF. We start by showing inductively that under the given assumptions, the statement

$$H_n: \quad \mathbb{I}_{A_n} T^n \mathbb{I}_{\mathcal{R}^c} = \mathbb{I}_{A_n} \quad \text{and} \quad (\forall a \in A_{n+1}^c) T \mathbb{I}_{A_n}(a) < 1 \quad \text{and} \quad (\forall a \in A_n^c) T^n \mathbb{I}_{\mathcal{R}^c}(a) < 1$$

holds for all $n \geq 0$. We first prove that the statement H_n holds for $n = 0$. The first and third statements of H_0 hold trivially. For the second statement, we have to prove that $T\mathbb{I}_{A_0}(a) < 1$ for all $a \in A_1^c = A_0^c \cup (A_0 \setminus A_1)$. On $A_0 \setminus A_1$, the desired inequality holds by

definition. On $A_0^c = \mathcal{R}$ it holds because there $\mathbb{T}\mathbb{I}_{A_0}$ is zero: no state in the top class \mathcal{R} has access to any state outside it.

Next, we prove that $H_n \Rightarrow H_{n+1}$. First of all,

$$\mathbf{T}^{n+1}\mathbb{I}_{A_0} = \mathbf{T}(\mathbf{T}^n\mathbb{I}_{A_0}) = \mathbf{T}[\mathbb{I}_{A_n}\mathbf{T}^n\mathbb{I}_{A_0} + \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] = \mathbf{T}[\mathbb{I}_{A_n} + \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}], \quad (5)$$

where the last equality follows from the induction hypothesis H_n . It follows from the definition of A_{n+1} that $\mathbb{I}_{A_{n+1}}\mathbf{T}\mathbb{I}_{A_n} = \mathbb{I}_{A_{n+1}}$, and therefore

$$\begin{aligned} \mathbb{I}_{A_{n+1}} &= \mathbb{I}_{A_{n+1}}\mathbf{T}[\mathbb{I}_{A_n} + \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0} - \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] \leq \mathbb{I}_{A_{n+1}}\mathbf{T}[\mathbb{I}_{A_n} + \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] + \mathbb{I}_{A_{n+1}}\mathbf{T}[-\mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] \\ &= \mathbb{I}_{A_{n+1}}\mathbf{T}^{n+1}\mathbb{I}_{A_0} + \mathbb{I}_{A_{n+1}}\mathbf{T}[-\mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] \\ &\leq \mathbb{I}_{A_{n+1}}\mathbf{T}^{n+1}\mathbb{I}_{A_0} \leq \mathbb{I}_{A_{n+1}}, \end{aligned}$$

where the first inequality follows from T2, the second from the fact that $-\mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0} \leq 0$ and therefore $\mathbb{I}_{A_{n+1}}\mathbf{T}[-\mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] \leq 0$ [use T1 and T5], and the third from $\mathbf{T}^{n+1}\mathbb{I}_{A_0} \leq 1$ [use T5]. The second equality follows from Eq. (5). Hence indeed $\mathbb{I}_{A_{n+1}} = \mathbb{I}_{A_{n+1}}\mathbf{T}^{n+1}\mathbb{I}_{A_0}$.

Next, observe that $A_{n+2}^c = A_{n+1}^c \cup (A_{n+1} \setminus A_{n+2})$. By definition, $\mathbb{T}\mathbb{I}_{A_{n+1}}(a) < 1$ for all $a \in A_{n+1} \setminus A_{n+2}$. It also follows from the induction hypothesis H_n that $\mathbb{T}\mathbb{I}_{A_n}(a) < 1$ for all $a \in A_{n+1}^c$. But since $A_{n+1} \subseteq A_n$, it follows from T5 that $\mathbb{T}\mathbb{I}_{A_{n+1}} \leq \mathbb{T}\mathbb{I}_{A_n}$, and therefore also $\mathbb{T}\mathbb{I}_{A_{n+1}}(a) < 1$ for all $a \in A_{n+1}^c$. Hence indeed $\mathbb{T}\mathbb{I}_{A_{n+1}}(a) < 1$ for all $a \in A_{n+2}^c$.

To finish the induction proof, let $\beta := \max_{a \in A_n^c} \mathbf{T}^n\mathbb{I}_{\mathcal{R}^c}(a)$, then $\beta < 1$ by the induction hypothesis H_n . We then infer from Eq. (5) that

$$\mathbf{T}^{n+1}\mathbb{I}_{A_0} = \mathbf{T}[\mathbb{I}_{A_n} + \mathbb{I}_{A_n^c}\mathbf{T}^n\mathbb{I}_{A_0}] \leq \mathbf{T}[\mathbb{I}_{A_n} + \beta\mathbb{I}_{A_n^c}] = \mathbf{T}[\beta + (1 - \beta)\mathbb{I}_{A_n}] = \beta + (1 - \beta)\mathbf{T}\mathbb{I}_{A_n}.$$

Consider any $a \in A_{n+1}^c$, then $\mathbb{T}\mathbb{I}_{A_n}(a) < 1$ by the induction hypothesis H_n , and therefore $\mathbf{T}^{n+1}\mathbb{I}_{A_0}(a) \leq \beta + (1 - \beta)\mathbb{T}\mathbb{I}_{A_n}(a) < 1$ since also $\beta < 1$. We conclude that H_{n+1} holds too.

To continue the proof, we observe that $A_0, A_1, \dots, A_n, \dots$ is a non-increasing sequence, and that A_0 is finite. This implies that there must be some first $k \in \mathbb{N}$ such that $A_{k+1} = A_k$. Clearly, $k \leq |A_0|$. We now prove by induction that $G_n : \mathbb{I}_{A_k}\mathbf{T}^{n+k}\mathbb{I}_{A_0} = \mathbb{I}_{A_k}$ for all $n \geq 0$. The statement G_n clearly holds for $n = 0$: it follows directly from H_k . We show that $G_n \Rightarrow G_{n+1}$. First of all,

$$\mathbf{T}^{n+k+1}\mathbb{I}_{A_0} = \mathbf{T}(\mathbf{T}^{n+k}\mathbb{I}_{A_0}) = \mathbf{T}[\mathbb{I}_{A_k}\mathbf{T}^{n+k}\mathbb{I}_{A_0} + \mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] = \mathbf{T}[\mathbb{I}_{A_k} + \mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}],$$

where the last equality follows from the induction hypothesis G_n . As before, it follows from the definition of A_{k+1} that $\mathbb{I}_{A_{k+1}}\mathbf{T}\mathbb{I}_{A_k} = \mathbb{I}_{A_{k+1}}$, and therefore $\mathbb{I}_{A_k}\mathbf{T}\mathbb{I}_{A_k} = \mathbb{I}_{A_k}$, so

$$\begin{aligned} \mathbb{I}_{A_k} &= \mathbb{I}_{A_k}\mathbf{T}[\mathbb{I}_{A_k} + \mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0} - \mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] \\ &\leq \mathbb{I}_{A_k}\mathbf{T}[\mathbb{I}_{A_k} + \mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] + \mathbb{I}_{A_k}\mathbf{T}[-\mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] \\ &= \mathbb{I}_{A_k}\mathbf{T}^{n+k+1}\mathbb{I}_{A_0} + \mathbb{I}_{A_k}\mathbf{T}[-\mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] \leq \mathbb{I}_{A_k}\mathbf{T}^{n+k+1}\mathbb{I}_{A_0} \leq \mathbb{I}_{A_k}, \end{aligned}$$

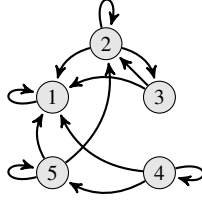
where the first inequality follows from T2, the second from the fact that $-\mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0} \leq 0$ and therefore $\mathbb{I}_{A_k}\mathbf{T}[-\mathbb{I}_{A_k^c}\mathbf{T}^{n+k}\mathbb{I}_{A_0}] \leq 0$ [use T1 and T5], and the third from $\mathbf{T}^{n+k+1}\mathbb{I}_{A_0} \leq 1$ [use T5]. Hence indeed $\mathbb{I}_{A_k} = \mathbb{I}_{A_k}\mathbf{T}^{n+k+1}\mathbb{I}_{A_0}$.

There are now two possibilities. The first is that $A_k \neq \emptyset$. It follows from the arguments above that for any element a of A_k , $T^\ell \mathbb{I}_{\mathcal{R}^c}(a) = 1$ for all $\ell \in \mathbb{N}$, which implies that T cannot be top class absorbing. The second possibility is that $A_k = \emptyset$. It follows from the argument above that $T^k \mathbb{I}_{\mathcal{R}^c}(a) < 1$ for all $a \in A_k^c = \mathcal{X}$ which implies that T is top class absorbing.

Example 3. Define $Tf = \max \{Mf : L \leq M \leq U \text{ and } M \text{ stochastic}\}$ where L and U are given by

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 0 & 0 \\ 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/4 \end{pmatrix} \text{ and } U = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 3/4 & 1/2 & 0 & 0 \\ 3/4 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1/4 & 3/4 & 0 & 0 & 1/4 \end{pmatrix}.$$

The corresponding upper accessibility graph $\overline{\mathcal{G}}(T)$ is given by



where $\{x\}$ corresponds to the unique strongly connected component that is final. As it is a singleton, it has cyclicity one, so there is a regular top class $\mathcal{R} = \{x\}$.

To check for top class absorption, we start iterating:

step 1: $T\mathbb{I}_{\mathcal{R}^c} = (0 \ 1 \ 1/2 \ 1 \ 1)^T$ whence $\mathbb{I}_{A_1} = (0 \ 1 \ 0 \ 1 \ 1)^T$,

step 2: $T\mathbb{I}_{A_1} = (0 \ 3/4 \ 1/2 \ 1 \ 1)^T$ whence $\mathbb{I}_{A_2} = (0 \ 0 \ 0 \ 1 \ 1)^T$,

step 3: $T\mathbb{I}_{A_2} = (0 \ 0 \ 0 \ 1 \ 1/4)^T$ whence $\mathbb{I}_{A_3} = (0 \ 0 \ 0 \ 1 \ 0)^T$,

step 4: $T\mathbb{I}_{A_3} = (0 \ 0 \ 0 \ 1 \ 0)^T$ whence $\mathbb{I}_{A_4} = (0 \ 0 \ 0 \ 1 \ 0)^T$.

Because $A_4 = A_3 \neq \emptyset$ we conclude that T is not top class absorbing and therefore not ergodic.

6. Coefficient of ergodicity

It is clear that ergodicity would follow immediately from Banach's fixed point theorem if T were contractive instead of non-expansive. With this in mind, one might think that conditions for ergodicity might coincide with contractiveness of T . This is not true. Take, for example, the particular upper transition operator $T = \mathbb{I}_{\mathcal{X}} \max$, which is not contractive, but, by Proposition 2, clearly ergodic.

In addition to requiring the sequence $\{T^k f\}$ to converge, ergodicity also requires that $\max T^n f - \min T^n f \rightarrow 0$ when $n \rightarrow \infty$. It seems therefore to be more natural to focus on the so-called *variation pseudo-norm* defined by:

$$\|f\|_v := \max f - \min f.$$

Under this pseudo-norm, upper transition operators will again be non-expansive. The extra condition that makes the map T contractive is expressed by Škulj and Hable [12] in terms of the coefficient of ergodicity. It is a standard trick, see of Nussbaum's monograph [7], to use Hilbert's projective metric to show contraction. The variation norm we define now can be seen as additive version of Hilbert's projective metric.

Proposition 7. *If we define the coefficient of ergodicity of an upper transition operator T as*

$$\rho(T) := \max \{\|Th\|_v : 0 \leq h \leq 1\}, \quad (6)$$

then T is ergodic if $\rho(T^m) < 1$ for some $m \in \mathbb{N}$.

PROOF. Consider any $f \in \mathcal{L}(\mathcal{X}^c)$. It follows by repeatedly applying T5, T3 and T4 that for all $k \in \mathbb{N}$:

$$\min f \leq \min T^k f \leq \min T^{k+1} f \leq \max T^{k+1} f \leq \max T^k f \leq \max f. \quad (7)$$

This tells us that the sequence $\min T^k f$ is non-decreasing and bounded above. It therefore converges to some real number m . Similarly, the sequence $\max T^k f$ is non-increasing and bounded below, and therefore converges to some real number M . It is also clear from Eq. (7) that $m \leq M$. Suppose that there is some $p \in \mathbb{N}$ such that $\rho(T^p) < 1$. Then we have to prove that $m = M$, which is what we now set out to do.

Since $0 \leq (f - \min f)/\|f\|_v \leq 1$, we infer from Eq. (6), T3 and T4 that

$$\frac{\|Tf\|_v}{\|f\|_v} = \|T \frac{f - \min f}{\|f\|_v}\|_v \leq \rho(T),$$

and therefore also

$$\|T^k f\|_v \leq \rho(T^k) \|f\|_v \text{ for all } k \in \mathbb{N}. \quad (8)$$

Then applying Eq. (8) repeatedly tells us that for the upper transition operator $\Lambda := T^p$:

$$\|\Lambda^n f\|_v \leq \rho(T^p)^n \|f\|_v \text{ for all } n \in \mathbb{N}.$$

But this implies that $\max \Lambda_n f - \min \Lambda_n f = \|\Lambda^n f\|_v \rightarrow 0$. Since we know from the arguments above that $\max \Lambda_n f \rightarrow M$ and $\min \Lambda_n f \rightarrow m$, this implies that indeed $m = M$.

Not only does the coefficient of ergodicity allow us to decide in favour of ergodicity, by Eq.(8) it also gives a numerical bound on the speed of convergence. The main problem however is that, in the worst case, in order to check for ergodicity in this manner, we need to calculate the coefficient of ergodicity of T^k for powers k up to $g(|\mathcal{X}^c|)$, where g is Landau's function. This renders this approach impractical from a computational point of view, making our approach preferable.

7. The critical graph versus the upper accessibility graph

The aim of Akian and Gaubert's paper [1] is to determine, for convex, monotone and non-expansive maps Φ , combinatorial bounds on orbit lengths of the described maps. Although the scope of Akian and Gaubert's paper is different, it essentially overlaps with this work. Akian and Gaubert try to describe the entire (additive) eigenspace of the map Φ . Their tool of choice for doing that is what they call the *critical graph* $\mathcal{G}^c(\Phi)$ of the map Φ . It is defined as the final graph $\mathcal{G}^f(\partial\Phi(v))$ of the subdifferential $\partial\Phi$ of Φ evaluated in an (additive) eigenvector v . Akian and Gaubert define the subdifferential of the operator Φ evaluated in any vector v as

$$\partial\Phi(v) := \left\{ M \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|} : (\forall f \in \mathbb{R}^{|\mathcal{X}|}) \Phi f - \Phi v \geq M(f - v) \right\}.$$

They show that the matrices M that belong to $\partial\Phi(v)$ are necessarily stochastic matrices.

Let us now consider what happens in the special case that Φ is an upper transition operator T , in order to better understand the relationship between their approach and ours. Given the constant additivity of T we can choose any constant vector as an (additive) eigenvector to calculate the critical graph. To make things as simple as possible, we opt for the zero gamble. The subdifferential of T evaluated in this eigenvector then becomes

$$\partial T(0) = \left\{ M \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|} : (\forall f \in \mathbb{R}^{|\mathcal{X}|}) Tf \geq M(f) \right\} = \mathcal{T},$$

which is the closed convex set of transition matrices that corresponds with the upper transition operator T , as defined by Eq. (1). The critical graph $\mathcal{G}^c(T) = \mathcal{G}^f(\partial T(0)) = \mathcal{G}^f(\mathcal{T})$ is then (defined as) the union of all the final graphs of the stochastic matrices belonging to \mathcal{T} . A final graph of a stochastic matrix can be found by interpreting this stochastic matrix as an adjacency matrix and restricting the corresponding graph to its final classes (see also the discussion in Sec. 5.1).

By comparing the definitions of the upper accessibility graph $\overline{\mathcal{G}}(T)$ and the critical graph $\mathcal{G}^c(T)$ for an upper transition operator T , we see that the strongly connected components of $\overline{\mathcal{G}}(T)$ have to be unions of strongly connected components of $\mathcal{G}^c(T)$. It is also not too difficult to see that the final classes of $\overline{\mathcal{G}}(T)$ and the final classes of $\mathcal{G}^c(T)$ are the same. This is exactly what allows us to check for top class regularity using the (usually much) cruder upper accessibility graph.

If the convex closed set of transition matrices \mathcal{T} corresponding with T is given explicitly in terms of a finite set of extreme points, then the calculation of the critical graph might be preferred over the calculation of the accessibility graph. However, if no finite set of extreme points is given, a vertex enumeration step is required (assuming that, unlike in Ex. 2, \mathcal{T} has a finite number of extreme points). As it is provable that any algorithm based on vertex enumeration cannot have polynomial time complexity, the algorithm given by Akian and Gaubert becomes in this case computationally intractable. This is where our algorithm stands out. The reason it does, is because it works directly with the upper transition operator, and drops extra eigenspace information that is not needed when checking for ergodicity.

8. Conclusion

In this paper we have given different equivalent conditions under which an upper transition operator—which corresponds to a set of non-stationary Markov chains—is ergodic. We have shown that ergodicity is completely determined by the eigenvalues and functions of the transition operator as is the case in classical Markov chains. This opens the door to a spectral theorem for upper transition operators. Unfortunately, it is at this point not known how to calculate these eigenvalues in general. This is why we developed an alternative test for ergodicity, which needs at most $2|\mathcal{X}| - 1$ evaluations of the upper transition operator. Any algorithm that implements this test consists of two steps: the first checks for top class regularity by building the upper accessibility graph and checking for final strongly connected components and their cyclicity. In some cases a second step is needed, to check for top class absorption.

Another approach that has been documented in the literature [12], calculates the coefficient of ergodicity and checks whether there is some power of the transition operator such that the corresponding coefficient becomes strictly smaller than one. If this is the case, then the non-expansive map that every upper transition operator is, becomes a contractive map and ergodicity is a fact. Interesting about the coefficient of ergodicity is that it moreover provides an upper bound on the speed of convergence. What makes this approach difficult to use outside the theoretical context, is that there is at present no efficient algorithm to calculate the coefficient of ergodicity. It is moreover likely that very high powers of the upper transition operator need to be calculated.

A paper with a different background is the very general work of Akian and Gauber [1], who describe an algorithm for checking ergodicity of upper transition operators that are piecewise affine. In practise, their algorithm relies heavily on extreme points to calculate the subdifferential. If the set of extreme points is given, then their critical graph approach is the shortest way to get to all qualitative information available on the eigenspace of the upper transition operator. If these extreme points are not given explicitly, than a vertex enumeration step is involved which is computationally very hard as any algorithm based on vertex enumeration cannot be polynomial time.

Our algorithm avoids the vertex enumeration step by using the upper transition operator directly. It also allows checking for ergodicity for upper transition operators whose ‘credal set’ has an infinite amount of extreme points. Of course, extra information about the eigenspace available through the critical graph approach, not necessary for deciding upon ergodicity, may be lost by using our simpler approach based on accessibility alone.

In a number of stochastic control applications that provide a motivation for Akian and Gaubert’s work [1], the extreme points of the polytopes of transition probability measures cannot be enumerated (only separation or minimisation oracles are available), and hence, dealing with such situations as we explain here, is also quite relevant in that application context.

Acknowledgements

This paper presents research results of the Belgian programme on Inter-university Attraction Poles, initiated by the Belgian Federal Science Policy Office. The scientific

responsibility rests with its authors. We are grateful to a number of anonymous reviewers for their many relevant and helpful comments, and for additional pointers to the literature.

- [1] Akian, M., Gaubert, S., JAN 2003. Spectral theorem for convex monotone homogeneous maps, and ergodic control. *Nonlinear Analysis-theory Methods & Applications* 52 (2), 637–679.
- [2] de Cooman, G., Hermans, F., Quaeghebeur, E., 2009. Imprecise markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences* 23 (4), 597–635.
- [3] Gaubert, S., Gunawardena, J., 2004. The perron-frobenius theorem for homogeneous, monotone functions. *Transaction of the Americal Mathematical Society* 356 (12), 4931–4950.
- [4] Hartfiel, D. J., 1998. Markov Set-Chains. No. 1695 in *Lecture Notes in Mathematics*. Springer, Berlin.
- [5] Kemeny, J. G., Snell, J. L., 1976. *Finite Markov Chains*. Undergraduate Text in Mathematics. Springer-Verlag, New York, reprint of the 1960 Edition.
- [6] Lemmens, B., Scheutzwow, M., JUN 2005. On the dynamics of sup-norm non-expansive maps. *Ergodic Theory and Dynamical Systems* 25 (Part 3), 861–871.
- [7] Nussbaum, R., SEP 1988. Hilberts projective metric and iterated nonlinear maps. *Memoirs of the American Mathematical Society* 75 (391), 1–&.
- [8] Nussbaum, R., 1990. Omega limit sets of nonexpansive maps: finiteness and cardinality estimates. *Differential Integral Equations* 3 (3), 523–540.
- [9] Nussbaum, R., 1991. Estimates of the periods of periodic points for nonexpansive operators. *Israel Journal of Mathematics* 76, 345–380, 10.1007/BF02773870. URL <http://dx.doi.org/10.1007/BF02773870>
- [10] Seneta, E., January 2006. *Non-negative Matrices and Markov Chains* (Springer Series in Statistics). Springer.
- [11] Sine, R., JUN 1990. A nonlinear perron-frobenius theorem. *Proceedings of the American Mathematical Society* 109 (2), 331–336.
- [12] Škulj, D., Hable, R., 2009. Coefficients of ergodicity for imprecise markov chains. In: Augustin, T., Coolen, F. P., Serafin, M., Troffaes, M. C. (Eds.), *ISIPTA '09 – Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*. SIPTA, pp. 377–386.
- [13] Walley, P., 1991. *Statistical Reasoning with Imprecise Probabilities*. Vol. 42 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.