Probability in the Engineering and Informational Sciences, **23**, 2009, 597–635. Printed in the U.S.A.
doi:10.1017/S0269964809990039

IMPRECISE MARKOV CHAINS AND THEIR LIMIT BEHAVIOR

Gert de Cooman, Filip Hermans, and Erik Quaeghebeur
SYSTeMS Research Group
Ghent University
Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium
E-mail: {gert.decooman, filip.hermans, erik.quaeghebeur}@ugent.be

When the initial and transition probabilities of a finite Markov chain in discrete time are not well known, we should perform a sensitivity analysis. This can be done by considering as basic uncertainty models the so-called *credal sets* that these probabilities are known or believed to belong to and by allowing the probabilities to vary over such sets. This leads to the definition of an *imprecise Markov chain*. We show that the time evolution of such a system can be studied very efficiently using so-called *lower* and *upper expectations*, which are equivalent mathematical representations of credal sets. We also study how the inferred credal set about the state at time n evolves as $n \to \infty$: under quite unrestrictive conditions, it converges to a uniquely invariant credal set, regardless of the credal set given for the initial state. This leads to a nontrivial generalization of the classical Perron–Frobenius theorem to imprecise Markov chains.

1. INTRODUCTION

One convenient way to model uncertain dynamical systems is to describe them as Markov chains. These have been studied in great detail, and their properties are well known. However, in many practical situations, it remains a challenge to accurately identify the transition probabilities in the Markov chain: The available information about physical systems is often imprecise and uncertain. Describing a real-life dynamical system as a Markov chain will therefore often involve unwarranted precision and might lead to conclusions not supported by the available information.

For this reason, it seems quite useful to perform probabilistic robustness studies, or sensitivity analyses, for Markov chains. This is especially relevant in decision-making applications. Many researchers in Markov Chain Decision Making

 0269-9648/09 $25.00

[12,18,25,36]—inspired by Satia and Lave's [27] original work—have paid attention to this issue of "imprecision" in Markov chains.

Work on the more mathematical aspects of modeling such imprecision in Markov chains was initiated in the early 1980s by Hartfiel and Senata (see [13–15]), under the name "Markov set-chains." Hartfiel's work seems to have been unknown to Kozine and Utkin [21], who approached the subject from a different angle. Armed with linear programming techniques, these authors performed an experimental study of the limit behavior of Markov chains with uncertain transition probabilities. More recently, Škulj [31,32] has also contributed to a formal study of the time evolution and limit behavior of such systems. Markov set-chains can also be seen as special cases of so-called *credal networks* under strong independence [4,5].

All of these approaches use *sets of probabilities* to deal with the imprecision in the transition probabilities. When these probabilities are not well known, they are assumed to belong to certain sets, and robustness analyses are performed by allowing the transition probabilities to vary over such sets. This should be contrasted with more common ways of performing a sensitivity analysis: looking at small deviations from a reference model and evaluating derivatives of important variables in this reference point.

As we will see, the sets of probabilities approach leads to a number of computational difficulties. However we will show that they can be overcome by tackling the problem from another angle, using lower and upper expectations rather than sets of probabilities. Our new method also makes it fairly easy to formulate and prove convergence (or Perron–Frobenius-like) results for Markov chains with uncertain transition probabilities that hold under weaker conditions than the ones found by Hartfiel [13,14] and Škulj [32]. We will see that our condition for this convergence, which requires that the imprecise Markov chain should be *regularly absorbing*, is implied by, and even strictly weaker than, both Hartfiel's *product scrambling* and Škulj's *regularity* conditions.

In the rest of this section, we give an overview of the theory of classical Markov chains and formulate the classical Perron–Frobenius theorem. Then in Sections 2 and 3 we introduce imprecise Markov chains and generalize many aspects of the classical theory. In Section 4 we briefly discuss accessibility relations, which allows us to give a nice interpretation to a number of conditions that will turn out to be sufficient for a Perron–Frobenius-like convergence result. In Section 5 we generalize the classical Perron–Frobenius theorem and explore the relation of our generalization with previous work in the literature. We discuss a number of theoretical and numerical examples in Section 6, and we give perspectives for further research in Section 7. Proofs of theorems and propositions have been relegated to the Appendix.

1.1. A Short Analysis of Classical Markov Chains

Consider a finite Markov chain in discrete time, where at consecutive times $n = 1, 2, 3, \dots, N$, $N \in \mathbb{N}$, the *state* $X(n)$ of a system can assume any value in a finite set $\mathcal{X}$. Here $\mathbb{N}$ denotes the set of nonzero natural numbers and N is the time horizon. The time evolution of such a system can be modeled as if it traversed a path in a so-called

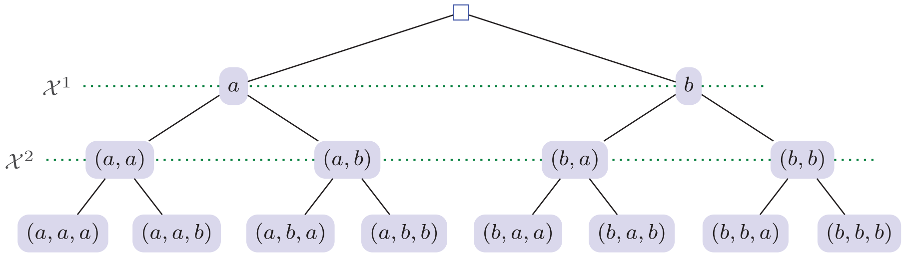

FIGURE 1. (Color online) The event tree for the time evolution of system that can be in two states, a and b, and can change state at time instants $n = 1, 2$. Also depicted are the respective cuts $\mathcal{X}^1$ and $\mathcal{X}^2$ of $\square$, where the states at times 1 and 2 are revealed.

event tree; see Shafer [29]. An example of such a tree for $\mathcal{X} = \{a, b\}$ and $N = 3$ is given in Figure 1.

The *situations*, or nodes, of the tree have the form $x_{1:k} := (x_1, \ldots, x_k) \in \mathcal{X}^k$, $k = 0, 1, \ldots, N$. For $k = 0$, there is some abuse of notation as we let $\mathcal{X}^0 := \{\square\}$, where $\square$ is the so-called *initial situation*, or root of the tree. In the cuts[1] $\mathcal{X}^n$ of $\square$, the value of the state $X(n)$ at time n is revealed.

In a classical analysis, it is generally assumed that we have (i) a probability distribution over the initial state $X(1)$, in the form of a probability mass function m_1 on $\mathcal{X}$, and (ii) for each situation $x_{1:n}$ that the system can be in at time n, a probability distribution over the next state $X(n+1)$, in the form of a probability mass function $q(\cdot|x_{1:n})$ on $\mathcal{X}$. This means that in each nonterminal situation[2] $x_{1:n}$ of the event tree, we have a *local* probability model telling us about the probabilities of each of its child nodes. This turns the event tree into a so-called *probability tree*; see Shafer [29, Chap. 3] and Kemeny and Snell [19, Sect. 1.9].

The probability tree for a Markov chain is special, because the *Markov condition* states that when the system jumps from state $X(n) = x_n$ to a new state $X(n+1)$, where the system goes to will only depend on the state $X(n) = x_n$ that the system was in at time n, not on its states $X(k) = x_k$ at previous times $k = 1, 2, \ldots, n-1$. In other words,

$$q(\cdot|x_{1:n}) = q_n(\cdot|x_n), \qquad x_{1:n} \in \mathcal{X}^n,\ n = 1, \ldots, N-1, \tag{1}$$

where $q_n(\cdot|x_n)$ is some probability mass function on $\mathcal{X}$. The Markov chain may be nonstationary, as the transition probabilities on the right-hand side in Eq. (1) are allowed to depend explicitly on the time n. Figure 2 gives an example of a probability tree for a Markov chain with $\mathcal{X} = \{a, b\}$ and $N = 3$.

With the local probability mass functions m_1 and $q_n(\cdot|x_n)$ we associate the linear real-valued *expectation functionals* E_1 and $E_n(\cdot|x_n)$, given, for all real-valued maps h on $\mathcal{X}$, by

$$E_1(h) := \sum_{x_1 \in \mathcal{X}} h(x_1) m_1(x_1) \quad \text{and} \quad E_n(h|x_n) := \sum_{x_{n+1} \in \mathcal{X}} h(x_{n+1}) q_n(x_{n+1}|x_n). \tag{2}$$

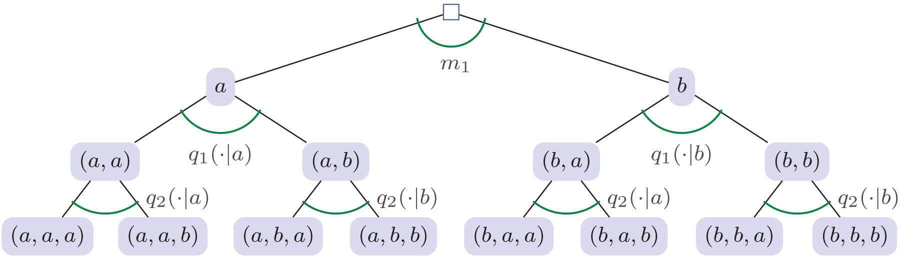

FIGURE 2. (Color online) The probability tree for the time evolution of a Markov chain that can be in two states, a and b, and can change state at each time instant $n = 1, 2$.

Throughout, we will formulate our results using expectations, rather than probabilities.[3] Our reasons for doing so are not merely aesthetic or a matter of personal preference; they will become clear as we go along.

In any probability tree, probabilities and expectations can be calculated very efficiently using backward recursion.[4] Suppose that in situation $x_{1:n}$, we want to calculate the conditional expectation $E(f|x_{1:n})$ of some real-valued map f on $\mathcal{X}^N$ that might depend on the values of the states $X(1), \ldots, X(N)$. Let us indicate briefly how this is done, also taking into account the simplifications due to the Markov condition (1).

For these simplifications, a prominent part will be played by the so-called *transition operators*[5] T_n and $\mathbb{T}_n$. Consider the linear space $\mathcal{L}(\mathcal{X})$ of all real-valued maps on $\mathcal{X}$. Then the linear operator (transformation) $\mathrm{T}_n\colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ is defined by

$$\mathrm{T}_n h(x_n) := E_n(h|x_n) = \sum_{x_{n+1}\in\mathcal{X}} h(x_{n+1}) q_n(x_{n+1}|x_n) \tag{3}$$

for all real-valued maps h on $\mathcal{X}$. In other words, $\mathrm{T}_n h$ is the real-valued map on $\mathcal{X}$ whose value $\mathrm{T}_n h(x_n)$ in $x_n \in \mathcal{X}$ is the conditional expectation of the random variable $h(X(n+1))$, given that the system is in state x_n at time n. More generally, we also consider the linear maps $\mathbb{T}_n$ from $\mathcal{L}(\mathcal{X}^{n+1})$ to $\mathcal{L}(\mathcal{X}^n)$, defined by

$$\begin{aligned}\mathbb{T}_n f(x_{1:n}) &:= \mathrm{T}_n(f(x_{1:n},\cdot))(x_n)\\ &= E_n(f(x_{1:n},\cdot)|x_n) = \sum_{x_{n+1}\in\mathcal{X}} f(x_{1:n}, x_{n+1}) q_n(x_{n+1}|x_n)\end{aligned} \tag{4}$$

for all $x_{1:n} \in \mathcal{X}^n$ and all real-valued maps f on $\mathcal{X}^{n+1}$.[6]

We begin our illustration of backward recursion by calculating $E(f|x_{1:n})$ for the case $n = N-1$. Here

$$\begin{aligned}E(f|x_{1:N-1}) &= E(f(x_{1:N-1},\cdot)|x_{1:N-1})\\ &= \sum_{x_N\in\mathcal{X}} f(x_{1:N-1}, x_N) q(x_N|x_{1:N-1})\end{aligned}$$

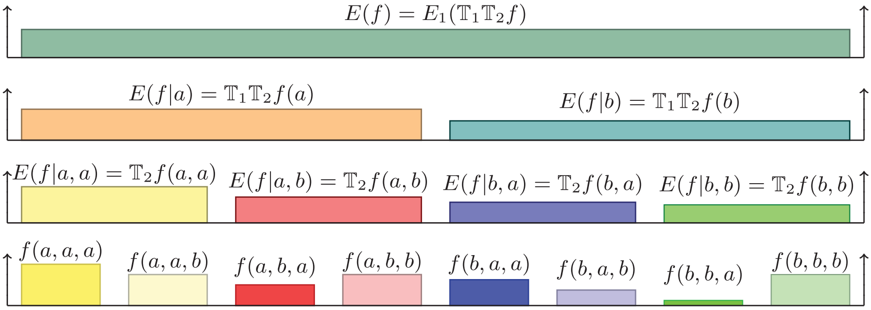

FIGURE 3. (Color online) Backward calculation of the conditional and joint expectations of a real-valued map f on $\mathcal{X}^3$, for a stationary Markov chain with state set $\mathcal{X} = \{a, b\}$, and a uniform probability mass function attached to each nonterminal situation.

$$= \sum_{x_N \in \mathcal{X}} f(x_{1:N-1}, x_N) q_{N-1}(x_N | x_{N-1}) = \mathbb{T}_{N-1} f(x_{1:N-1}), \tag{5}$$

where the third inequality follows from the Markov condition (1) and the fourth follows from Eq. (4). Using similar arguments for $n = N - 2$, we derive from the Law of Iterated Expectations[7] that

$$E(f|x_{1:N-2}) = E(E(f(x_{1:N-2}, \cdot, \cdot)|x_{1:N-2}, \cdot)|x_{1:N-2}) = \mathbb{T}_{N-2}\mathbb{T}_{N-1} f(x_{1:N-2}). \tag{6}$$

Repeating this argument leads to the backward recursion formulas

$$E(f|x_{1:n}) = \mathbb{T}_n \mathbb{T}_{n+1} \cdots \mathbb{T}_{N-1} f(x_{1:n}) \tag{7}$$

for $n = 1, \ldots, N - 1$, whereas for $n = 0$, we get

$$E(f) := E(f|\square) = E_1(\mathbb{T}_1 \mathbb{T}_2 \cdots \mathbb{T}_{N-1} f). \tag{8}$$

In these formulas, f is any real-valued map on $\mathcal{X}^N$. In Figure 3, we give a graphical representation of calculations using the backward recursion formulas (7) and (8), for a two-state stationary Markov chain.

For instance, if we let f be the indicator functions $I_{\{x_{1:N}\}}$ of the singletons $\{x_{1:N}\}$, formulas (7) and (8) allow us to calculate the joint probability mass function p defined by $p(x_{1:N}) = E(I_{\{x_{1:N}\}})$ for all the variables $X(1), \ldots, X(N)$. We can also use them to find the conditional mass functions $p_n(\cdot|x_n)$ and $p(\cdot|x_{1:n})$ defined by $p_n(x_{n+1:N}|x_n) = p(x_{n+1:N}|x_{1:n}) = E(I_{\{x_{1:N}\}}|x_{1:n})$.

1.2. The Perron–Frobenius Theorem for Classical Markov Chains

We are especially interested in the case of a *stationary* Markov chain and in the (marginal) expectation $E_n(h)$ of a real-valued map h (on $\mathcal{X}$) that depends only on the state $X(n)$ at time n. Here, Eq. (8) becomes

$$E_n(h) := E_1(\mathrm{T}^{n-1}h), \tag{9}$$

where $\mathrm{T} := \mathrm{T}_1 = \mathrm{T}_2 = \cdots = \mathrm{T}_{N-1}$ and where we denote by T^k the k-fold composition of T with itself; in particular, T^0 is the identity operator id on $\mathcal{L}(\mathcal{X})$. If we let $h = I_{\{x_n\}}$, this allows us to find the probability mass function $m_n(x_n) = E_n(I_{\{x_n\}})$, $x_n \in \mathcal{X}$ for the state $X(n)$.

By the way, the linear transition operator T is very closely related to the so-called *Markov*, or *transition*, *matrix* T of the stationary Markov chain, whose elements for all $(x, y) \in \mathcal{X}^2$ are defined by

$$T_{xy} := q(y|x) = \mathrm{T}I_{\{y\}}(x). \tag{10}$$

Any such transition matrix satisfies the conditions $T_{xy} \geq 0$ and $\sum_{z\in\mathcal{X}} T_{xz} = 1$. We will henceforth call *transition matrix* any matrix satisfying these properties.[8] The probability counterpart of the expectation formula (9) can then be written in matrix form as

$$m_n = m_1 T^{n-1}, \tag{11}$$

where, here and further on, we also use the notation m_n for the row vector whose components are the probabilities $m_n(x_n)$, $x_n \in \mathcal{X}$.

Under some restrictions on the transition operator T, the classical Perron–Frobenius theorem then tells us that as n (as well as the time horizon N) recedes to infinity, this probability mass function m_n converges to some limit, independently of the initial probability mass function m_1; see Kemeny and snell [19, Them. 4.1.6] and Luenberger [22, Chap. 6]. In terms of expectation functionals and transition operators, we have Theorem 1.1.

THEOREM 1.1 (Classical Perron–Frobenius Theorem, Expectation Form): *Consider a stationary Markov chain with finite state set $\mathcal{X}$ and transition operator* T. *Suppose that* T *is regular, meaning that there is some $k > 0$ such that* $\min \mathrm{T}^k I_{\{x\}} > 0$ *for all x in $\mathcal{X}$.*[9] *Then for every initial expectation operator E_1, the expectation operator $E_n = E_1 \circ \mathrm{T}^{n-1}$ for the state at time n converges pointwise to the same limit expectation operator E_∞:*

$$\lim_{n\to\infty} E_n(h) = \lim_{n\to\infty} E_1(\mathrm{T}^{n-1}h) =: E_\infty(h) \quad \textit{for all } h \in \mathcal{L}(\mathcal{X}). \tag{12}$$

Moreover, the limit expectation E_∞ is the only T-*invariant expectation on $\mathcal{L}(\mathcal{X})$, in the sense that $E_\infty = E_\infty \circ \mathrm{T}$.*

2. TOWARD IMPRECISE MARKOV CHAINS

The above treatment rests on the assumption that the initial probabilities and the transition probabilities are precisely known. If such is not the case, then it seems necessary to perform some kind of sensitivity analysis, in order to find out to what extent any conclusions we might reach using such a treatment depend on the actual values of these probabilities.

A very general way of performing a sensitivity analysis for probabilities involves calculations with closed convex sets of probability mass functions, also called *credal sets*, rather than with single probability measures. Let $\Sigma_{\mathcal{X}}$ denote the set of all probability mass functions on $\mathcal{X}$, an $(|\mathcal{X}|-1)$-dimensional unit simplex in the $|\mathcal{X}|$-dimensional linear space $\mathbb{R}^{\mathcal{X}}$; then $\{m \in \Sigma_{\mathcal{X}} \colon (\forall x \in \mathcal{X}) m(x) \leq 1/2\}$ is a credal set but $\{m \in \Sigma_{\mathcal{X}} \colon (\exists x \in \mathcal{X}) m(x) \geq 1/2\}$ is not.

There is a growing body of literature on this interesting and fairly new area of *imprecise probabilities*, starting with the publication of Walley's [33] seminal work. We refer to the literature [7,33–35] for more details and discussion.

Let us recall a number of results for credal sets, important for the developments in this article. Proofs can be found in Walley's book [33, Chaps. 2 and 3]. Specifying a closed convex set $\mathcal{P}$ of probability mass functions p on a finite set $\mathcal{Y}$ is equivalent to specifying its *lower* and *upper expectation* (functionals) $\underline{E}_{\mathcal{P}} \colon \mathcal{L}(\mathcal{Y}) \to \mathbb{R}$ and $\overline{E}_{\mathcal{P}} \colon \mathcal{L}(\mathcal{Y}) \to \mathbb{R}$, respectively, defined for all $g \in \mathcal{L}(\mathcal{Y})$ by

$$\underline{E}_{\mathcal{P}}(g) := \min \left\{E_p(g) \colon p \in \mathcal{P}\right\} \quad \text{and} \quad \overline{E}_{\mathcal{P}}(g) := \max \left\{E_p(g) \colon p \in \mathcal{P}\right\}, \tag{13}$$

where $E_p(g) = \sum_{y \in \mathcal{Y}} g(y)p(y)$ is the expectation of g associated with the probability mass function p. In a sensitivity analysis, such functionals are quite useful because they give tight lower and upper bounds on the expectation of any real-valued map. Since the functionals $\underline{E}_{\mathcal{P}}$ and $\overline{E}_{\mathcal{P}}$ are *conjugate* in the sense that $\underline{E}_{\mathcal{P}}(g) = -\overline{E}_{\mathcal{P}}(-g)$ for all real-valued maps g on $\mathcal{Y}$, one is completely determined if the other is known. Below, we concentrate on upper expectations. Any upper expectation $\overline{E} = \overline{E}_{\mathcal{P}}$ associated with some credal set $\mathcal{P}$ satisfies the following properties (see, e.g., [33, Sect. 2.6.1]):

- $\overline{E}1$. $\min g \leq \overline{E}(g) \leq \max g$ for all g in $\mathcal{L}(\mathcal{Y})$ (boundedness);
- $\overline{E}2$. $\overline{E}(g_1 + g_2) \leq \overline{E}(g_1) + \overline{E}(g_2)$ for all g_1 and g_2 in $\mathcal{L}(\mathcal{Y})$ (subadditivity);
- $\overline{E}3$. $\overline{E}(\lambda g) = \lambda \overline{E}(g)$ for all real $\lambda \geq 0$ and all g in $\mathcal{L}(\mathcal{Y})$ (nonnegative homogeneity);
- $\overline{E}4$. $\overline{E}(g + \mu I_{\mathcal{X}}) = \overline{E}(g) + \mu$ for all real μ and all g in $\mathcal{L}(\mathcal{Y})$ (constant additivity);
- $\overline{E}5$. If $g_1 \leq g_2$, then $\overline{E}(g_1) \leq \overline{E}(g_2)$ for all g_1 and g_2 in $\mathcal{L}(\mathcal{Y})$ (monotonicity);
- $\overline{E}6$. If $g_n \to g$ pointwise, then $\overline{E}(g_n) \to \overline{E}(g)$ for all sequences g_n in $\mathcal{L}(\mathcal{Y})$ (continuity);
- $\overline{E}7$. $\overline{E}(g) \geq -\overline{E}(-g) = \underline{E}(g)$ for all g in $\mathcal{L}(\mathcal{Y})$ (upper–lower consistency).

Conversely, for any real functional $\overline{E}$ that is defined on $\mathcal{L}(\mathcal{Y})$ and that satisfies the conditions $\overline{E}1$–$\overline{E}3$, there is a unique credal set $\mathcal{P} \subseteq \Sigma_{\mathcal{X}}$ such that $\overline{E}$ coincides with the upper expectation $\overline{E}_{\mathcal{P}}$; namely, $\mathcal{P} = \left\{p \in \Sigma_{\mathcal{Y}} \colon (\forall f \in \mathcal{L}(\mathcal{Y})) E_p(f) \leq \overline{E}(f)\right\}$. Such an $\overline{E}$ therefore automatically also satisfies conditions $\overline{E}4$–$\overline{E}7$. It therefore make sense to *call* upper expectation *any real functional $\overline{E}$ on $\mathcal{L}(\mathcal{Y})$ that satisfies properties $\overline{E}1$–$\overline{E}3$.*

What is the upshot of all this for the Markov chain problem we are considering here? First, in the initial situation $\square$, corresponding to time $n = 0$, rather than a single initial probability mass function m_1, we now have a local credal set $\mathcal{M}_1$ of candidate mass functions m_1 for the state $X(1)$ that the system will be in at time $k = 1$. We denote by $\overline{E}_1$ the upper expectation associated with $\mathcal{M}_1$:

$$\overline{E}_1(h) := \max\left\{\sum_{x\in\mathcal{X}} h(x)m_1(x) : m_1 \in \mathcal{M}_1\right\} \quad \text{for all } h \in \mathcal{L}(\mathcal{X}). \tag{14}$$

Additionally, in any situation $x_{1:n} \in \mathcal{X}^n$ corresponding to time $n = 1, 2, \ldots, N-1$, instead of a single transition mass function $q_n(\cdot|x_n)$, we now have a local credal set $\mathcal{Q}_n(\cdot|x_n)$ of candidate conditional mass functions $q_n(\cdot|x_n)$ for the state $X(n+1)$ that the system will be in at time $n+1$. We denote by $\overline{E}_n(\cdot|x_n)$ the upper expectation associated with $\mathcal{Q}_n(\cdot|x_n)$; that is,

$$\overline{E}_n(h|x_n) := \max\left\{\sum_{x\in\mathcal{X}} h(x)q(x) : q \in \mathcal{Q}_n(\cdot|x_n)\right\} \quad \text{for all } h \in \mathcal{L}(\mathcal{X}). \tag{15}$$

We call the resulting model an *imprecise Markov chain*. Figure 4 gives an example of a probability tree for an imprecise Markov chain. It is an imprecise-probability tree where the local conditional models satisfy the *Markov condition*:

$$\mathcal{Q}(\cdot|x_{1:n}) = \mathcal{Q}(\cdot|x_n) \quad \text{for all } x_{1:n} \in \mathcal{X}^n \text{ and } n = 1, 2, \ldots, N-1. \tag{16}$$

A classical, or *precise*, Markov chain is an imprecise one with credal sets that are singletons.

How, then, can a sensitivity analysis be performed for such an imprecise Markov chain? We choose, in each nonterminal situation $x_{1:k}$ of the above-mentioned event tree, a local transition probability mass $q(\cdot|x_{1:k})$ in the set of possible candidates $\mathcal{Q}_k(\cdot|x_k)$.[10] For $k = 0$, we get the initial situation $\square$, where we choose some element m_1 in the set of possible candidates $\mathcal{M}_1$. By making a choice of local model for each nonterminal

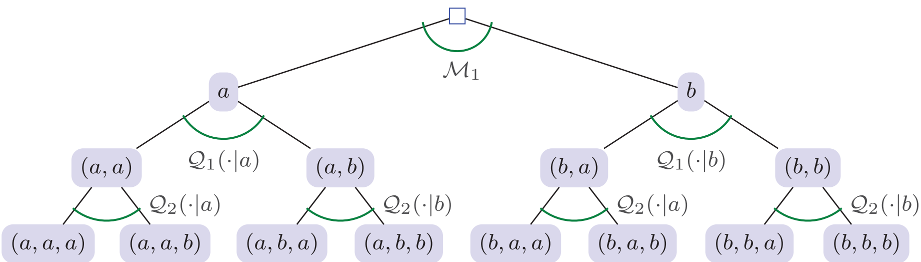

FIGURE 4. (Color online) The tree for the time evolution of an imprecise Markov chain that can be in two states, a and b, and can change state at each time instant $n = 1, 2$.

situation in the event tree, we obtain what we call a *compatible probability tree*, for which we may calculate all (conditional) expectations and probability mass functions:

$$E(f|x_{1:n}) = \sum_{x_{n+1:N}\in\mathcal{X}^{N-n}} f(x_{1:n}, x_{n+1:N}) \prod_{k=n}^{N-1} q(x_{k+1}|x_{1:k}), \tag{17}$$

$$E(f) = \sum_{x_{1:N}\in\mathcal{X}^{N}} f(x_{1:N}) m_1(x_1) \prod_{k=1}^{N-1} q(x_{k+1}|x_{1:k}), \tag{18}$$

for $n = 1, \ldots, N-1$ and for all real-valued maps f on $\mathcal{X}^N$. As we have just come to realize, the probability trees that are compatible with an imprecise Markov chain are no longer necessarily (precise) Markov chains themselves. It is still possible to calculate the $E(f|x_{1:n})$ and $E(f)$ in Eqs (17) and (18) using backward recursion [29, Chap. 3], but the formulas for doing so will be more complicated than the ones for precise Markov chains given by Eqs. (7) and (8).

If we repeat this for every other choice of the m_1 in $\mathcal{M}_1$ and the $q(\cdot|x_{1:k})$ in $\mathcal{Q}_k(\cdot|x_k)$, we end up with an infinity of compatible probability trees (except when all the credal sets are singleton, of course), for which the associated (conditional) expectations and probability mass functions turn out to constitute closed convex sets. We denote their corresponding upper expectation functionals on $\mathcal{L}(\mathcal{X}^N)$ by $\overline{E}(\cdot|x_{1:n})$ and $\overline{E}$. These upper expectations, and the conjugate lower expectations, are the final aim of our sensitivity analysis.

The procedure we have just described is computationally very complex. When the closed convex sets $\mathcal{M}_1$ and $\mathcal{Q}_k(\cdot|x)$ each have a finite number of extreme points (polytopes), we can limit ourselves to working with these sets of extreme points rather than with the infinite sets themselves. However even then, the computational complexity of this approach will generally be exponential in the number of time steps.

However, we will see in Section 3 that the upper expectations $\overline{E}$ and $\overline{E}(\cdot|x_{1:n})$ associated with the closed convex sets of (conditional) probability mass functions for the compatible probability trees of an imprecise Markov chain can be calculated in the same way as the expectations E and $E(\cdot|x_{1:n})$ in a precise one: using counterparts of the backward recursion formulas (7)–(9). Because of this, making inferences about the mass function of the state at time n, (i.e., finding the upper envelope $\overline{E}_n$ of the E_n given in Eq. (9)) *now has a complexity that is linear, rather than exponential, in the number of time steps n.* This is our first contribution.

Our second contribution in this article is a Perron–Frobenius theorem for a special class of so-called regularly absorbing stationary imprecise Markov chains: in Section 5 we prove a generalization of Theorem 1.1, which tells us that under fairly weak conditions, the upper expectation operators $\overline{E}_n$ converge to limits that do not depend on the initial upper expectation operators $\overline{E}_1$. Our result also extends a number of other related convergence theorems for imprecise Markov chains in the literature [13–15,32].

3. SENSITIVITY ANALYSIS OF IMPRECISE MARKOV CHAINS

We can now take our most important step: deriving the backward recursion formulas for the conditional and joint upper expectations in an imprecise Markov chain. We first define *upper transition operators* $\overline{\mathrm{T}}_n$ and $\overline{\mathbb{T}}_n$. The operator $\overline{\mathrm{T}}_n \colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ is defined by

$$\overline{\mathrm{T}}_n h(x_n) := \overline{E}_n(h|x_n) \tag{19}$$

for all real-valued maps h on $\mathcal{X}$ and all x_n in $\mathcal{X}$. In other words, $\overline{\mathrm{T}}_n h$ is the real-valued map on $\mathcal{X}$, whose value $\overline{\mathrm{T}}_n h(x_n)$ in $x_n \in \mathcal{X}$ is the conditional upper expectation of the random variable $h(X(n+1))$, given that the system is in state x_n at time n. More generally, we also consider the maps $\overline{\mathbb{T}}_n$ from $\mathcal{L}(\mathcal{X}^{n+1})$ to $\mathcal{L}(\mathcal{X}^n)$, defined by

$$\overline{\mathbb{T}}_n f(x_{1:n}) := \big(\overline{\mathrm{T}}_n f(x_{1:n}, \cdot)\big)(x_n) = \overline{E}_n(f(x_{1:n}, \cdot)|x_n) \tag{20}$$

for all $x_{1:n}$ in $\mathcal{X}^n$ and all real-valued maps f on $\mathcal{X}^{n+1}$. Of course, we can also consider lower expectations and lower transition operators, which are related to the upper expectations and upper transition operators by conjugacy. As is the case for upper expectations, it is possible to introduce the notion of an upper transition operator directly, by basing it on a number of defining properties rather than by referring to an underlying imprecise Markov chain. We refer to the Appendix for more details.

The upper expectations $\overline{E}(\cdot|x_{1:n})$ and $\overline{E}$ on $\mathcal{L}(\mathcal{X}^N)$ can be calculated very easily by backwards recursion, cfr. (7) and (8).

THEOREM 3.1 (Concatenation Formula): *For any $x_{1:n}$ in $\mathcal{X}^n$, $n = 1, \dots, N-1$ and for any real-valued map f on $\mathcal{X}^N$,*

$$\overline{E}(f|x_{1:n}) = \overline{\mathbb{T}}_n \overline{\mathbb{T}}_{n+1} \cdots \overline{\mathbb{T}}_{N-1} f(x_{1:n}), \tag{21}$$

$$\overline{E}(f) = \overline{E}_1(\overline{\mathbb{T}}_1 \overline{\mathbb{T}}_2 \cdots \overline{\mathbb{T}}_{N-1} f). \tag{22}$$

Call, for any nonempty subset I of $\{1, \dots, N\}$, a real-valued map f on $\mathcal{X}^N$ *I-measurable* if $f(x_{1:N}) = f(z_{1:N})$ for all $x_{1:N}$ and $z_{1:N}$ in $\mathcal{X}^N$ such that $x_k = z_k$ for all $k \in I$. In other words, an I-measurable f only depends on the states $X(k)$ at times $k \in I$. As an example, an $\{n\}$-measurable map h only depends on the state $X(n)$ at time n, and we identify it with a map on $\mathcal{X}$ (but remember that it acts on states at time n). The following proposition tells us that all conditional upper expectations satisfy a Markov condition (cf. (1)).

PROPOSITION 3.2 (Markov Condition): *Consider an imprecise Markov chain with finite state set $\mathcal{X}$ and time horizon N. Fix $n \in \{1, \dots, N-1\}$. Let $x_{1:n-1}$ and $z_{1:n-1}$ be arbitrary elements of $\mathcal{X}^{n-1}$ and let $x_n \in \mathcal{X}$. Let f be any $\{n, n+1, \dots, N\}$-measurable real-valued map on $\mathcal{X}^N$. Then $\overline{E}(f|x_{1:n-1}, x_n) = \overline{E}(f|z_{1:n-1}, x_n)$, so we can write $\overline{E}(f|x_{1:n-1}, x_n) = \overline{E}_{|n}(f|x_n)$.*

The index $|n$ is intended to make clear that we are considering an expectation conditional on the state $X(n)$ at time n.

If we apply the joint upper expectation $\overline{E}$ to maps h that only depend on the state $X(n)$ at time n, we get the *marginal upper expectation* $\overline{E}_n(h) := \overline{E}(h)$, and $\overline{E}_n$ is a model for the uncertainty about the state $X(n)$ at time n. More generally, taking into account Proposition 3.2, we use the notation $\overline{E}_{n|\ell}(h|x_\ell) := \overline{E}_{|\ell}(h|x_\ell)$ for the upper expectation of $h(X(n))$, conditional on $X(\ell) = x_\ell$ with $1 \leq \ell < n$. With notations established in Eq. (15), $\overline{E}_{n+1|n}(h|x_n) = \overline{E}_n(h|x_n) = \overline{\mathrm{T}}_n h(x_n)$. Such expectations can be found using simpler recursion formulas than Eqs (21) and (22), as they are based on the simpler upper transition operators $\overline{\mathrm{T}}_k$.

COROLLARY 3.3: *For any real-valued map h on $\mathcal{X}$ and for any $1 \leq \ell < n \leq N$ and all x_ℓ in $\mathcal{X}$;*

$$\overline{E}_{n|\ell}(h|x_\ell) = \overline{\mathrm{T}}_\ell \overline{\mathrm{T}}_{\ell+1} \cdots \overline{\mathrm{T}}_{n-1} h(x_\ell) \quad \text{and} \quad \overline{E}_n(h) = \overline{E}_1(\overline{\mathrm{T}}_1 \overline{\mathrm{T}}_2 \cdots \overline{\mathrm{T}}_{n-1} h). \qquad \textbf{(23)}$$

This offers a reason for formulating our theory in terms of real-valued maps rather than events: Suppose we want to calculate the upper probability $\overline{E}_n(A)$ that the state $X(n)$ at time n belongs to the set A. According to Eq. (23), $\overline{E}_n(A) = \overline{E}_1(\overline{\mathrm{T}}_1 \cdots \overline{\mathrm{T}}_{n-1} I_A)$, and even if $\overline{\mathrm{T}}_{n-1} I_A$ can still be calculated using upper probabilities only, it will generally assume values other than zero and 1 and, therefore, will generally not be the indicator of some event. Already after one step (i.e., in order to calculate $\overline{\mathrm{T}}_{n-2} \overline{\mathrm{T}}_{n-1} I_A$), we need to leave the ambit of events, and turn to the more general real-valued maps, even if we only want to calculate upper *probabilities* after n steps.

For joint upper and lower probability mass functions, however, we can remain within the ambit of events.

PROPOSITION 3.4 (Chapman–Kolmogorov Equations): *For an imprecise Markov chain, we have for all $1 \leq n < m \leq N$ and all $(x_n, x_{n+1:m}) \in \mathcal{X}^{m-n+1}$ that*

$$\overline{E}_{|n}(\{x_{n+1:m}\}|x_n) = \prod_{k=n}^{m-1} \overline{\mathrm{T}}_k I_{\{x_{k+1}\}}(x_k), \qquad \textbf{(24)}$$

and for all $1 \leq m \leq N$ and all $x_{1:m} \in \mathcal{X}^m$ we have that

$$\overline{E}(\{x_{1:m}\}) = \overline{E}_1(\{x_1\}) \prod_{k=1}^{m-1} \overline{\mathrm{T}}_k I_{\{x_{k+1}\}}(x_k). \qquad \textbf{(25)}$$

There are analogous expressions for the lower expectations.

4. ACCESSIBILITY RELATIONS

From now on, and for the rest of the article, we mainly consider *stationary imprecise Markov chains with an infinite time horizon*. This means that for each time $n \in \mathbb{N}$, we consider the same upper transition operator $\overline{\mathrm{T}}_n = \overline{\mathrm{T}}$.

The classification of the states of such a stationary (im)precise Markov chain can be fruitfully started by introducing a so-called *accessibility relation* $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$: Let x and y be any two states in $\mathcal{X}$ and let n be a number of steps in $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, then $x \overset{n}{\rightsquigarrow} y$ expresses that y is accessible from x in n steps. To be an accessibility relation, a generic ternary relation $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ has to satisfy the following defining properties:

$$(\forall x, y \in \mathcal{X})(x \overset{0}{\rightsquigarrow} y \Leftrightarrow x = y), \tag{26}$$

$$(\forall x, y, z \in \mathcal{X})(\forall m, n \in \mathbb{N}_0)(x \overset{n}{\rightsquigarrow} y \quad \text{and} \quad y \overset{m}{\rightsquigarrow} z \Rightarrow x \overset{n+m}{\rightsquigarrow} z), \tag{27}$$

$$(\forall x \in \mathcal{X})(\forall n \in \mathbb{N})(\exists y \in \mathcal{X}) x \overset{n}{\rightsquigarrow} y. \tag{28}$$

An accessibility relation is classically derived from the transition matrix of a stationary Markov chain; in Section 4.2 we will associate such a relation with a stationary imprecise Markov chain. However, for *any* (abstract) accessibility relation satisfying the conditions (26)–(28), we can draw all of the following conclusions, regardless of what transition matrix or operator it was derived from or whether it comes about in any other way; Kemeny and Snell [19, Sect. 1.4] gave a detailed justification. In what follows, we use the terminology introduced by Kemeny and Snell, but we want to remind the reader that the terms we use might also have various other meanings in different parts of the literature.

4.1. Abstract Accessibility Relations

Accessibility relations give rise to many interesting concepts, which we discuss next. We refer to Figure 5 for a graphical representation.

Consider any two states x and y in $\mathcal{X}$. Then y is *accessible from* x, which we denote as $x \rightsquigarrow y$, if there is some $n \in \mathbb{N}_0$ such that $x \overset{n}{\rightsquigarrow} y$. If x and y are accessible from one another, then we say that x and y *communicate*, which we denote as $x \leftrightsquigarrow y$.

It follows from Eqs (26) and (27) that the binary relation $\rightsquigarrow$ on $\mathcal{X}$ is a preorder (i.e., is reflexive and transitive). The binary relation $\leftrightsquigarrow$ on $\mathcal{X}$ is the associated equivalence relation. This *communication relation* $\leftrightsquigarrow$ partitions the state set $\mathcal{X}$ into equivalence classes D of states that are accessible from one another, called *communication classes*. The preorder $\rightsquigarrow$ induces a partial order on this partition, also denoted by $\rightsquigarrow$.

Undominated or *maximal* states with respect to the preorder $\rightsquigarrow$ are states x such that $x \rightsquigarrow y \Rightarrow y \rightsquigarrow x$ for any state y in $\mathcal{X}$. This means that a maximal state has access only to other maximal states in the same communication class, and to no other states. Collections of maximal states, such as the communication classes they belong to, are also called *maximal*. The other states and collections of them, such as the communication classes they belong to, are called *transient*. If all maximal states communicate, or in other words if there is a unique maximal communication class, this class is called the *top* class. It is made up of those states that are accessible from any state.

Consider, for any x and y in $\mathcal{X}$, the set

$$N_{xy} := \{n \in \mathbb{N} : x \overset{n}{\rightsquigarrow} y\}; \tag{29}$$

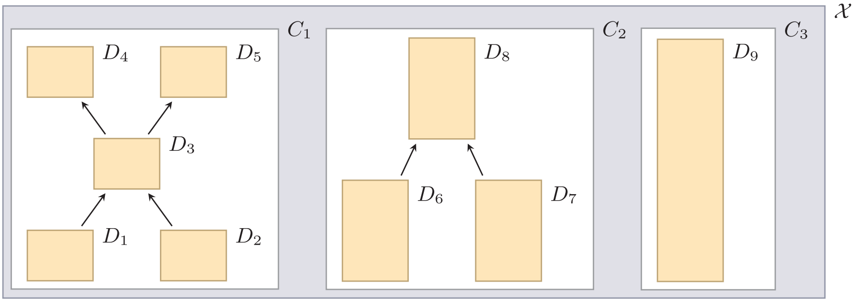

FIGURE 5. (Color online) Three increasingly finer partitions of the state set $\mathcal{X}$ for a particular stationary (im)precise Markov chain, or, more generally, for an accessibility relation $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$. No transition between states of the classes C_1, C_2, and C_3 is possible, and these classes can be seen as separate (im)precise Markov chains. The equivalence classes D_k for the communication relation are partially ordered by the relation $\rightsquigarrow$, whose (Hasse) diagram is represented by the upward arrows. Maximal classes are D_4, D_5, D_8, and D_9, the other classes are transient. If D_4, D_5, D_8, and D_9 are aperiodic, the accessibility relation restricted to respectively C_1, C_2, and C_3 is respectively maximal class regular, top class regular, and regular.

that is, those numbers of steps after which y is accessible from x. We call the *period* d_x of a state x the greatest common divisor of the set N_{xx} (i.e., $d_x := \gcd\{n \in \mathbb{N} : x \overset{n}{\rightsquigarrow} x\}$). Because, by Eq. (27), N_{xx} is closed under addition, we can rely on a basic number-theoretic result (see, e.g., Kemeny and Snell [19, Thm. 1.4.1]), which tells us that N_{xx} is, up to perhaps a finite number of initial elements, equal to the set of all multiples of d_x.

Now, consider an equivalence class D of communicating states and any two states x and y in that class. Then it is not difficult to show that they have the same period: $x \leftrightsquigarrow y \Rightarrow d_x = d_y$. We denote by d_D the common period of all elements of the equivalence class D.

PROPOSITION 4.1: *Consider arbitrary x and y in some maximal communication class D. Then there is some $0 \leq t_{xy} < d_D$ such that $n \in N_{xy} \Rightarrow n \equiv t_{xy} \pmod{d_D}$; that is, n and t_{xy} are equal up to some multiple of d_D. Moreover,*

$$(\exists n \in \mathbb{N})(\forall k \geq n)\, t_{xy} + kd_D \in N_{xy}. \tag{30}$$

For any x, y, and z in this equivalence class D, $t_{xy} + t_{yz} \equiv t_{xz} \pmod{d_D}$, and therefore $t_{yz} = 0$ if and only if $t_{xy} = t_{xz}$. This implies that $t_{yz} = 0$ determines an equivalence relation on this equivalence class D, which further partitions it into d_D subsets, called *cyclic classes*. In such a cyclic class, all states y give the same value to t_{xy}, for any given x in D. Within D, the system moves from cyclic class to cyclic class, in a definite ordered cycle of length d_D. If D is transient, then in some cyclic classes it is possible that, rather than moving to the next cyclic class, the system moves to (a state in) another equivalence class D' for the communication relation that is a successor to D for the partial order $\rightsquigarrow$.

If $d_D = 1$, or in other words if $t_{xy} = 0$ for all $x, y \in D$, then there is only one cyclic class in D, and we call the communication class D, and all its states, *aperiodic*. Moreover, if D is maximal, then D is called *regular*. The following general characterization of regularity is easily derived from Proposition 4.1; see also Kemeny and Snell's arguments [19, Chapts. 1 and 4].

PROPOSITION 4.2: *A communication class $D \subseteq \mathcal{X}$ is regular under the accessibility relation $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ if and only if*

$$(\exists n \in \mathbb{N})(\forall k \geq n)(\forall x, y \in D)x \overset{k}{\rightsquigarrow} y. \tag{31}$$

An interesting special case obtains when there is only one equivalence class for the communication relation (viz. $\mathcal{X}$), so $\mathcal{X}$ is maximal, and there is only one cyclic class (viz. $\mathcal{X}$), meaning that all states are aperiodic. In that case, the accessibility relation $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ is called *regular* as well. If all maximal communication classes are regular (aperiodic), the accessibility relation is called *maximal class regular*. If there is only one maximal communication class and if this top class is, moreover, regular (aperiodic), then the accessibility relation is called *top class regular*. Top class regularity has the following simple alternative characterization.

PROPOSITION 4.3: *An accessibility relation $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ is top class regular if and only if the corresponding set $\mathcal{R}_{\rightsquigarrow}$ of so-called* maximal regular states *is nonempty:*

$$\mathcal{R}_{\rightsquigarrow} = \{x \in \mathcal{X} \colon (\exists n \in \mathbb{N})(\forall k \geq n)(\forall y \in \mathcal{X})y \overset{k}{\rightsquigarrow} x\} \neq \emptyset; \tag{32}$$

in that case, this set $\mathcal{R}_{\rightsquigarrow}$ is the top communication class.

4.2. Accessibility Relations for Imprecise Markov Chains

Because we now only consider stationary imprecise Markov chains, this means that for each time $n \in \mathbb{N}$, we consider the same transition models $\mathcal{Q}_n(\cdot|x) = \mathcal{Q}(\cdot|x), x \in \mathcal{X}$ or, equivalently, for the upper transition operators: $\overline{\mathrm{T}}_n = \overline{\mathrm{T}}$ and $\overline{\mathbb{T}}_n = \overline{\mathbb{T}}$.

Let us denote by $\overline{P}_{xy}^n$ the upper probability of going in n steps from state x to state y. For $n = 0$, $\overline{P}_{xy}^0 = I_{\{y\}}(x)$, and for $n \geq 1$, $\overline{P}_{xy}^n = \overline{E}_{k+n|k}(\{y\}|x)$, where—because of stationarity—the right-hand side does not depend on $k \in \mathbb{N}$. By Corollary 3.3, we find that $\overline{P}_{xy}^n = \overline{\mathrm{T}}^n I_{\{y\}}(x)$ for all $n \in \mathbb{N}_0$. The following two propositions allow us to associate an accessibility relation with the upper transition operator. They are immediate generalizations of similar results involving (precise) probabilities in (precise) Markov chains.

PROPOSITION 4.4: *For all x, y, and z in $\mathcal{X}$ and for all n and m in $\mathbb{N}_0$,*

$$\overline{P}_{xy}^{n+m} \geq \overline{P}_{xz}^n \overline{P}_{zy}^m. \tag{33}$$

PROPOSITION 4.5: *For all x in $\mathcal{X}$ and for all n in $\mathbb{N}_0$, there is some y in $\mathcal{X}$ such that* $\overline{P}^n_{xy} > 0$.

Because of these results, which ensure that Eqs (27) and (28) are satisfied [Eq. (26) is trivially satisfied because $\overline{P}^0_{xy} = I_{\{y\}}(x)$], we can define an accessibility relation $\cdot \overset{\cdot}{\to} \cdot$ using $\overline{P}^n_{xy}$: for any x and y in $\mathcal{X}$ and any $n \in \mathbb{N}_0$,

$$x \overset{n}{\to} y \Leftrightarrow \overline{P}^n_{xy} > 0 \Leftrightarrow \overline{\mathrm{T}}^n I_{\{y\}}(x) > 0. \tag{34}$$

Clearly, $x \overset{n}{\to} y$ if there is *some* compatible probability tree in which it is possible (meaning that there is a nonzero probability) to go from state x to y in n time steps. In other words, $x \overset{n}{\to} y$ if it is not considered impossible in the context of our imprecise-probability model to go from x to y in n steps: We then say that y is *accessible* from x in n steps; and if $x \to y$, then y is *accessible* from x.

The following notion will be essential for the convergence result we present in Section 5. It involves both lower and upper transition probabilities.

DEFINITION 4.1 (Regularly Absorbing): *A stationary imprecise Markov chain is called* regularly absorbing *if it is top class regular (under $\to$), meaning that*

$$\mathcal{R}_{\to} := \left\{x \in \mathcal{X} \colon (\exists n \in \mathbb{N})(\forall k \geq n)(\forall y \in \mathcal{X})\overline{\mathrm{T}}^k I_{\{x\}}(y) > 0\right\} \neq \emptyset, \tag{35}$$

and if, moreover, for all y in $\mathcal{X} \setminus \mathcal{R}_{\to}$ there is some $n \in \mathbb{N}$ such that $\underline{\mathrm{T}}^n I_{\mathcal{R}_{\to}}(y) > 0$.

In particular, an imprecise Markov chain that is regular (under $\to$, meaning that the accessibility relation $\to$ is regular) is also regularly absorbing (under $\to$) in a trivial way.

5. CONVERGENCE FOR STATIONARY IMPRECISE MARKOV CHAINS

We call an upper expectation $\overline{E}$ on $\mathcal{L}(\mathcal{X})$ $\overline{\mathrm{T}}$-*invariant* whenever $\overline{E} \circ \overline{\mathrm{T}} = \overline{E}$, so whenever $\overline{E}(\overline{\mathrm{T}}h) = \overline{E}(h)$ for all $h \in \mathcal{L}(\mathcal{X})$.

THEOREM 5.1 (Perron–Frobenius Theorem, Upper Expectation Form): *Consider a stationary imprecise Markov chain with finite state set $\mathcal{X}$ that is regularly absorbing. Then for every initial upper expectation $\overline{E}_1$, the upper expectation $\overline{E}_n = \overline{E}_1 \circ \overline{\mathrm{T}}^{n-1}$ for the state at time n converges pointwise to the same upper expectation $\overline{E}_\infty$:*

$$\lim_{n\to\infty} \overline{E}_n(h) = \lim_{n\to\infty} \overline{E}_1(\overline{\mathrm{T}}^{n-1}h) =: \overline{E}_\infty(h) \quad \text{for all } h \text{ in } \mathcal{L}(\mathcal{X}). \tag{36}$$

Moreover, the limit upper expectation $\overline{E}_\infty$ is the only $\overline{\mathrm{T}}$-invariant upper expectation on $\mathcal{L}(\mathcal{X})$.

Let us compare this convergence result to what exists in the literature. The classical Perron–Frobenius theorem 1.1 is, of course, a special case of our Theorem 5.1, because if (the transition operator of) a precise stationary Markov chain is regular in the sense of Theorem 1.1, then it is also regular (under $\to$) and, therefore, regularly absorbing.

Other authors have presented convergence results for stationary imprecise Markov chains, namely Hartfiel and Seneta [15], Hartfiel [13,14], and Škulj [32]. They all use the following approach. They consider some set $\mathcal{T}$ of (one-step) transition matrixes T and deduce from that a corresponding set $\mathcal{T}^n$ of n-step transition matrixes given by

$$\mathcal{T}^n := \{T_1 T_2 \cdots T_n \colon T_1, T_2, \cdots, T_n \in \mathcal{T}\}. \tag{37}$$

Hartfiel called the sequence $\mathcal{T}^n$, $n \in \mathbb{N}$, a *Markov set chain*. If we also have a set $\mathcal{M}_1$ of (marginal) mass functions m_1 for $X(1)$, then they take the corresponding set $\mathcal{M}_n$ of (marginal) mass functions for $X(n)$ to be

$$\mathcal{M}_n = \left\{m_1 T \colon m_1 \in \mathcal{M}_1 \text{ and } T \in \mathcal{T}^{n-1}\right\}, \tag{38}$$

where, as earlier, we also denote by m the row vector corresponding to the mass function m. Furthermore, if we also denote by h the column vector corresponding to the values $h(x)$ of the real-valued map h in all $x \in \mathcal{X}$, then we find that the corresponding set $\mathcal{E}_n(h)$ of expectations of $h(X(n))$ is given by

$$\mathcal{E}_n(h) = \left\{m_1 T h \colon m_1 \in \mathcal{M}_1 \text{ and } T \in \mathcal{T}^{n-1}\right\}. \tag{39}$$

Incidentally, these are also the formulas that can be obtained by considering imprecise Markov chains to be special cases of so-called credal networks under a strong independence assumption; for more details, see Cozman's work [4,5] for instance.

Škulj [32] considered the set $\mathcal{T}$ of transition matrixes T corresponding to a so-called *interval stochastic matrix*, meaning that $\mathcal{T}$ is the set of all transition matrixes such that $\underline{T} \leq T \leq \overline{T}$, where $\underline{T}$ and $\overline{T}$ are so-called lower and upper transition matrixes; see also Section 6.3 for the related model in terms of upper transition operators. Hartfiel [13] considered arbitrary sets of transition matrixes, but in his book [14] he also focused mainly on interval stochastic matrixes.

What is the relationship between the Markov set-chain model and the model involving upper transition operators we have studied and motivated above? Consider a stationary imprecise Markov chain with upper transition operator $\overline{\mathrm{T}}$. For each state x, as $\overline{\mathrm{T}}h(x)$ has been defined as a conditional upper expectation $\overline{E}(h|x)$, there is a corresponding credal set $\mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$ given by

$$\mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x) := \left\{q(\cdot|x) \in \Sigma_{\mathcal{X}} \colon (\forall h \in \mathcal{L}(\mathcal{X})) E_{q(\cdot|x)}(h) \leq \overline{\mathrm{T}}h(x)\right\}, \tag{40}$$

and then also

$$\overline{\mathrm{T}}h(x) = \max\left\{E_{q(\cdot|x)}(h) \colon q(\cdot|x) \in \mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)\right\}. \tag{41}$$

With these credal sets, we can associate a set of transition matrixes $\mathcal{T}_{\overline{\mathrm{T}}}$:

$$\mathcal{T}_{\overline{\mathrm{T}}} := \left\{T \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}} \colon (\forall x \in \mathcal{X})(\exists q(\cdot|x) \in \mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x))(\forall y \in \mathcal{X}) T_{xy} = q(y|x)\right\}. \tag{42}$$

In other words, each row $T_{x\cdot}$ of any such transition matrix is formed by the transition probabilities corresponding to some element of $\mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$. The elements T of

$\mathcal{T}_{\overline{\mathrm{T}}}$ are the transition matrixes that can be constructed using the one-step information contained in the conditional credal sets $\mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$ and, therefore, in the (one-step) upper transition operator $\overline{\mathrm{T}}$. More generally, the set $\mathcal{T}_{\overline{\mathrm{T}}^n}$ contains all n-step transition matrixes that correspond to the n-step upper transition operator $\overline{\mathrm{T}}^n$ (see the Appendix for more details about why we can also consider $\overline{\mathrm{T}}^n$ to be an upper transition operator).

PROPOSITION 5.2: *Consider a stationary imprecise Markov chain with upper transition operator* $\overline{\mathrm{T}}$ *and let* $n \in \mathbb{N}$. *Then we have the following:*

(i) $\mathcal{T}_{\overline{\mathrm{T}}}^n \subseteq \mathcal{T}_{\overline{\mathrm{T}}^n}$.

(ii) For all real-valued maps h *on* $\mathcal{X}$, *there is some* $T \in \mathcal{T}_{\overline{\mathrm{T}}}^n$ *such that for all* $x \in \mathcal{X}$, $\overline{\mathrm{T}}^n h(x) = (Th)_x$.

(iii) For all real-valued maps h *on* $\mathcal{X}$ *and all* $x \in \mathcal{X}$,

$$\overline{\mathrm{T}}^n h(x) = \max\left\{(Th)_x \colon T \in \mathcal{T}_{\overline{\mathrm{T}}}^n\right\} \quad \textit{and} \quad \underline{\mathrm{T}}^n h(x) = \min\left\{(Th)_x \colon T \in \mathcal{T}_{\overline{\mathrm{T}}}^n\right\}. \tag{43}$$

We gather from the following counterexample that for $n > 1$, $\mathcal{T}_{\overline{\mathrm{T}}}^n$ can be strictly included in $\mathcal{T}_{\overline{\mathrm{T}}^n}$. This shows that the model based on imprecise-probability trees and upper transition operators that we have been using is more detailed than the Markov set-chain model. Nevertheless, as Proposition 5.2(iii) indicates, both models yield very strongly related (if not identical) results as far as the calculation of marginal expectations for $X(n)$ is concerned.

Example 5.1: Consider $\overline{\mathrm{T}} := (1-\varepsilon)\,\mathrm{id} + I_{\mathcal{X}}\varepsilon \max$, where $0 \leq \varepsilon \leq 1$ and id is the identity operator, which leaves its argument real-valued map h unchanged: $\mathrm{id}\, h = h$. This corresponds to a special case of the contamination models (47) discussed in Section 6.1. For the corresponding two-step transition operator, we find that $\overline{\mathrm{T}}^2 = (1-\delta)\,\mathrm{id} + I_{\mathcal{X}}\delta \max$, with $\delta := \varepsilon(2-\varepsilon)$.

Let $|\mathcal{X}| = 2$, then the sets of corresponding transition matrixes are

$$\mathcal{T}_{\overline{\mathrm{T}}} = \left\{\begin{bmatrix} 1-\varepsilon_1 & \varepsilon_1 \\ \varepsilon_2 & 1-\varepsilon_2 \end{bmatrix} : 0 \leq \varepsilon_1, \varepsilon_2 \leq \varepsilon\right\} \quad \text{and}$$

$$\mathcal{T}_{\overline{\mathrm{T}}^2} = \left\{\begin{bmatrix} 1-\delta_1 & \delta_1 \\ \delta_2 & 1-\delta_2 \end{bmatrix} : 0 \leq \delta_1, \delta_2 \leq \delta\right\}. \tag{44}$$

We now show that the set $\mathcal{T}_{\overline{\mathrm{T}}}^2$ is *strictly* contained in $\mathcal{T}_{\overline{\mathrm{T}}^2}$. Any element of $\mathcal{T}_{\overline{\mathrm{T}}}^2$ is given by

$$\begin{bmatrix} 1-\varepsilon_1 & \varepsilon_1 \\ \varepsilon_2 & 1-\varepsilon_2 \end{bmatrix}\begin{bmatrix} 1-\varepsilon_3 & \varepsilon_3 \\ \varepsilon_4 & 1-\varepsilon_4 \end{bmatrix}$$
$$= \begin{bmatrix} 1-\varepsilon_1-\varepsilon_3+\varepsilon_1\varepsilon_3+\varepsilon_1\varepsilon_4 & \varepsilon_1+\varepsilon_3-\varepsilon_1\varepsilon_3-\varepsilon_1\varepsilon_4 \\ \varepsilon_2+\varepsilon_4-\varepsilon_2\varepsilon_4-\varepsilon_2\varepsilon_3 & 1-\varepsilon_2-\varepsilon_4+\varepsilon_2\varepsilon_4+\varepsilon_2\varepsilon_3 \end{bmatrix} \tag{45}$$

for some $0 \leq \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \leq \varepsilon$ and therefore clearly belongs to $\mathcal{T}_{\overline{\mathrm{T}}^2}$. However, it is straightforward to check that no choice of $\varepsilon_1, \varepsilon_2, \varepsilon_3$, or ε_4 in $[0, \varepsilon]$ corresponds to the element of $\mathcal{T}_{\overline{\mathrm{T}}^2}$ with $\delta_1 = \delta_2 = \delta = \varepsilon(2 - \varepsilon)$.

Škulj [32] called a compact set $\mathcal{T}$ of transition matrices *regular* if there is some $n > 0$ such that $T_{xy} > 0$ for all $T \in \mathcal{T}^n$ and all $x, y \in \mathcal{X}$. He then showed that for such regular $\mathcal{T}$ and for all compact $\mathcal{M}_1$, the corresponding sequence of compact sets $\mathcal{M}_n$, converges in Hausdorff norm to the same compact (and invariant) set $\mathcal{M}_\infty$. It follows that for all h and all compact $\mathcal{M}_1$, the sequence of compact sets $\mathcal{E}_n(h)$ will converge to the same compact set $\mathcal{E}_\infty(h)$. This is a clear generalization of the classical Perron–Frobenius Theorem 1.1. However, it follows from Proposition 5.2 that for a given stationary imprecise Markov chain with upper transition operator $\overline{\mathrm{T}}$, the set $\mathcal{T}_{\overline{\mathrm{T}}}$ is regular in Škulj's sense if and only if for some $n \in \mathbb{N}$, $\underline{\mathrm{T}}^n I_{\{y\}}(x) > 0$ for all $x, y \in \mathcal{X}$. This is much stronger than even our strongest convergence requirement of regularity (under $\rightarrow$), which only involves the condition $\overline{\mathrm{T}}^n I_{\{y\}}(x) > 0$ for all $x, y \in \mathcal{X}$. Škulj also proved a convergence result for conservative (too large) approximations of the $\overline{E}_n$, in the special case of a regular (under $\rightarrow$) imprecise Markov chain whose upper transition operator is 2-alternating; see Section 6.3 for further details.

We now turn to Hartfiel's [13–15] results. The strongest general convergence result seems to appear in his book [14, Sect. 3.2], where he used the *coefficient of ergodicity* $\tau(T)$ of a transition matrix T, defined by

$$\tau(T) = \frac{1}{2} \max_{x,y \in \mathcal{X}} \sum_{z \in \mathcal{X}} |T_{xz} - T_{yz}| = 1 - \min_{x,y \in \mathcal{X}} \sum_{z \in \mathcal{X}} \min\{T_{xz}, T_{yz}\}. \tag{46}$$

A transition matrix is called *scrambling* if $\tau(T) < 1$. Hartfiel called a compact set $\mathcal{T}$ of transition matrices *product scrambling* if there is some $m \in \mathbb{N}$ such that $\tau(T) < 1$ for all $T \in \mathcal{T}^m$. He then showed that for such product scrambling $\mathcal{T}$ and for all compact $\mathcal{M}_1$, the corresponding sequence of compact sets $\mathcal{M}_n$ converges in Hausdorff norm to the same compact (and invariant) set $\mathcal{M}_\infty$. Again, this is a generalization of the classical Perron–Frobenius theorem, and it includes Škulj's above-mentioned result as a special case. We believe, however, that this approach, based on the coefficient of ergodicity, has a number of drawbacks that our treatment does not have: The condition seems quite hard to check in practice and it it is hard to interpret directly. We now also argue that it is too strong, at least from our point of view.

PROPOSITION 5.3: *Consider a stationary imprecise Markov chain with upper transition operator* $\overline{\mathrm{T}}$. *If* $\mathcal{T}_{\overline{\mathrm{T}}}$ *is product scrambling, then the chain is regularly absorbing.*

Moreover, as the following counterexample shows, it is easy to find examples of stationary imprecise Markov chains that are regularly absorbing but for which the corresponding set $\mathcal{T}_{\overline{\mathrm{T}}}$ is not product scrambling. Another, perhaps more involved, such counterexample will be presented near the end of Section 6.4.

Example 5.2 (Vacuous Imprecise Markov chain): Consider an arbitrary state set $\mathcal{X}$ with at least two elements, and the upper transition operator $\overline{\mathrm{T}}$ defined by $\overline{\mathrm{T}}h = I_{\mathcal{X}} \max h$ for all real-valued maps h on $\mathcal{X}$. The set $\mathcal{T}_{\overline{\mathrm{T}}}$ that corresponds to this upper transition operator is the set of *all* transition matrixes $\mathcal{T}_{\mathrm{all}}$ and, consequently $\mathcal{T}_{\overline{\mathrm{T}}^n} = \mathcal{T}_{\overline{\mathrm{T}}}^n = \mathcal{T}_{\mathrm{all}}$ for all $n \in \mathbb{N}$ as well.

Consider the unit transition matrix T defined by $T_{xy} = \delta_{xy}$ [Kronecker delta], so the system remains with probability 1 in any state x that it is in. This T belongs to $\mathcal{T}_{\overline{\mathrm{T}}^n} = \mathcal{T}_{\mathrm{all}}$ for all $n \in \mathbb{N}$, but $\tau(T) = 1$, so $\mathcal{T}_{\mathrm{all}}$ is not product scrambling.

However, the chain is regularly absorbing! It is even regular (under $\to$), in a trivial way: $\overline{\mathrm{T}}^n I_{\{y\}}(x) = 1$ for all $n \in \mathbb{N}$ and all $x, y \in \mathcal{X}$. Observe that $\overline{\mathrm{T}}^n = I_{\mathcal{X}} \max$ and, therefore, $\overline{E}_\infty = \max$ for all $\overline{E}_1$.

6. EXAMPLES

In this section, we indicate how the theory developed in the previous sections can be applied in a number of practical situations. For each of these, the upper expectations are of some special types that are described in the literature on imprecise probabilities. We present concrete and explicit examples, as well as a number of simulations.

6.1. Contamination Models

Suppose we consider a precise stationary Markov chain, with transition operator T. We contaminate it with a vacuous model; that is, we take a convex mixture with the upper transition operator $I_{\mathcal{X}} \max$ of Example 5.2. This leads to the upper transition operator $\overline{\mathrm{T}}$, defined by

$$\overline{\mathrm{T}}h = (1 - \varepsilon)\mathrm{T}h + I_{\mathcal{X}}\varepsilon \max h, \tag{47}$$

for all $h \in \mathcal{L}(\mathcal{X})$, where ε is some constant in the open real interval $(0, 1)$. The underlying idea is that we consider a specific convex neighborhood of T. Since for all x in $\mathcal{X}$, $\min \overline{\mathrm{T}} I_{\{x\}} = (1 - \varepsilon) \min \mathrm{T} I_{\{x\}} + \varepsilon > 0$, this upper transition operator (or the associated imprecise Markov chain) is always regular (under $\to$), regardless of whether T is regular (in the sense of Theorem 1.1)! We infer from Theorem 5.1 that whatever the initial upper expectation operator $\overline{E}_1$ is, the upper expectation operator $\overline{E}_n$ for the state $X(n)$ at time $n \in \mathbb{N}$ will always converge to the same $\overline{E}_\infty$.

What is this $\overline{E}_\infty$ for given T and ε? For any $n \geq 1$,

$$\overline{\mathrm{T}}^n h = (1 - \varepsilon)^n \mathrm{T}^n h + I_{\mathcal{X}}\varepsilon \sum_{k=0}^{n-1} (1 - \varepsilon)^k \max \mathrm{T}^k h \tag{48}$$

and, therefore,

$$\overline{E}_{n+1}(h) = (1 - \varepsilon)^n \overline{E}_1(\mathrm{T}^n h) + \varepsilon \sum_{k=0}^{n-1} (1 - \varepsilon)^k \max \mathrm{T}^k h. \tag{49}$$

If we now let $n \to \infty$, we see that the limit is indeed independent of the initial upper expectation $\overline{E}_1$:

$$\overline{E}_\infty(h) = \varepsilon \sum_{k=0}^{\infty} (1-\varepsilon)^k \max \mathrm{T}^k h. \tag{50}$$

Example 6.1 (Contaminating a Cycle): Consider, for instance, $\mathcal{X} = \{a, b\}$ and let the precise Markov chain be the cycle with period 2, with transition operator T given by $\mathrm{T}h(a) = h(b)$ and $\mathrm{T}h(b) = h(a)$. Then $\mathrm{T}^{2n}h = h$ and $\mathrm{T}^{2n+1}h = \mathrm{T}h$ and, therefore, $\max \mathrm{T}^{2n}h = \max \mathrm{T}^{2n+1}h = \max h$, from which $\overline{E}_\infty(h) = \max h$. So the limit upper expectation is vacuous: we lose all information about the value of $X(n)$ as $n \to \infty$.

Example 6.2 (Contaminating a Random Walk): Consider a random walk, where $\mathcal{X} = \{a, b\}$ and $\mathrm{T}h = I_{\mathcal{X}}(h(a) + h(b))/2$. Then we find that $\overline{E}_\infty(h) = \varepsilon \max h + (1 - \varepsilon)(h(a) + h(b))/2$.

Example 6.3 (Another Contamination Model): To illustrate the convergence properties of an imprecise Markov chain, let us look at a simple numerical example. Again consider $\mathcal{X} = \{a, b\}$ and let the stationary imprecise Markov chain be defined by an initial credal set $\mathcal{M}_1 = \left\{m \in \Sigma_{\{a,b\}} : 0.6 \leq m(a) \leq 0.9\right\}$ and a contamination model of the type (47), with $\varepsilon = 0.1$ and for which the precise transition operator T is defined by the transition matrix

$$T := \begin{bmatrix} q(a|a) & q(b|a) \\ q(a|b) & q(b|b) \end{bmatrix} = \begin{bmatrix} 0.15 & 0.85 \\ 0.85 & 0.15 \end{bmatrix}.$$

In Figure 6 we have plotted the evolution of $\overline{E}_n(\{a\})$ and $\underline{E}_n(\{a\})$, the upper and lower probability for finding the system in state a at time n, which can be calculated efficiently using Eq. (49).

For comparison, we have also plotted the evolution of $E_n(\{a\})$, the probability for finding the system in state a at time n, for a (precise) Markov chain defined by probability mass functions that lie on the boundaries of the credal sets defining the above imprecise Markov chain; to wit, its initial mass function is given by the row vector $m_1 := [m_1(a) \;\; m_1(b)] = [0.9 \;\; 0.1]$ and its transition matrix is $\left[\begin{smallmatrix} 0.135 & 0.865 \\ 0.865 & 0.135 \end{smallmatrix}\right]$. Here, $E_\infty(\{a\}) = E_\infty(\{b\}) = 0.5$.

6.2. Belief Function Models

The contamination models we have just described are a special case of a more general and quite interesting class of models, based on Shafer's [28] notion of a belief function. We can consider a number of subsets $F_j, j = 1, \ldots, n$ of $\mathcal{X}$, and a convex mixture of

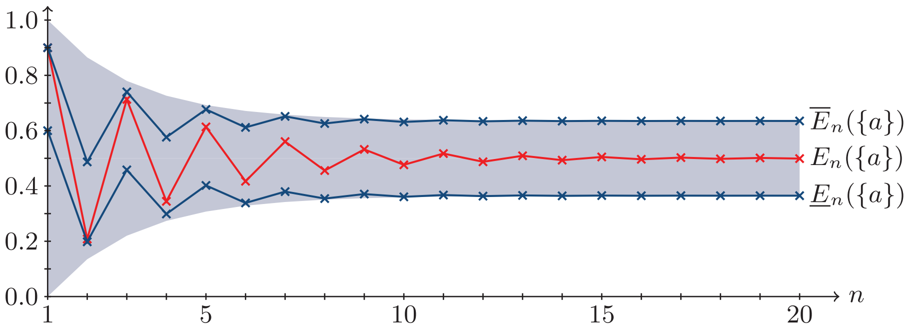

FIGURE 6. (Color online) The time evolution of (i) the upper and lower probability of finding the imprecise Markov chain of Example 6.3 in the state a (outer plot marks and connecting lines) and (ii) the probability of finding the classical Markov chain of Example 6.3 in the state a (inner plot marks and connecting lines). The filled area denotes the hull of the evolution of this probability, under the contamination model of Example 6.3, for all possible initial mass functions.

the vacuous upper expectations relative to these subsets:

$$\overline{E}(h) = \sum_{j=1}^{n} m(F_j) \max_{x \in F_j} h(x), \tag{51}$$

with $m(F_j) \geq 0$ and $\sum_{j=1}^{n} m(F_j) = 1$. In Shafer's terminology, the sets F_j are called *focal elements* and the $m(F_j)$'s are called the *basic probability assignment*.[11]

We can now consider imprecise Markov chains for which the local models, attached to the nonterminal situations in the tree, are of this type. The general backward recursion formulas we have given in Section 3 can then be used in combination with the simple formulas of the type (51) for an efficient calculation of all conditional and joint upper and lower expectations in the tree. We leave this implicit however and move on to another example, which is rather more popular in the literature.

6.3. Models With Lower and Upper Mass Functions

An intuitive way to introduce imprecise Markov chains [2,14,21,31] goes by way of so-called *probability intervals*, studied in a paper by de Campos, Huete, and Moral [3]; see also Walley [33, Sect. 4.6.1] and Hartfiel [14, Sect. 2.1]. It consists in specifying lower and upper bounds for mass functions. Let us explain how this is done in the specific context of Markov chains.

For the initial mass function m_1, we specify a lower bound $\underline{m}_1 : \mathcal{X} \to \mathbb{R}$, also called a *lower mass function*, and an upper bound $\overline{m}_1 : \mathcal{X} \to \mathbb{R}$, called an *upper mass function*. The credal set $\mathcal{M}_1$ attached to the initial situation, which corresponds to these bounds, is then given by

$$\mathcal{M}_1 := \left\{m \in \Sigma_{\mathcal{X}} \colon (\forall x \in \mathcal{X})\, \underline{m}_1(x) \leq m(x) \leq \overline{m}_1(x)\right\}. \tag{52}$$

Similarly, in each nonterminal situation $x_{1:k} \in \mathcal{X}^k$, $k = 1, \dots, N-1$, we have a credal set $\mathcal{Q}_k(\cdot|x_k)$ that is defined in terms of conditional lower and upper mass functions $\underline{q}_k(\cdot|x_k)$ and $\overline{q}_k(\cdot|x_k)$. Here, for instance, $\underline{q}_k(x_{k+1}|x_k)$ gives a lower bound on the transition probability $q_k(x_{k+1}|x_k)$ to go from state $X(k) = x_k$ to state $X(k+1) = x_{k+1}$ at time k.

Under some consistency conditions (for more details, see [3]), the upper expectation associated with $\mathcal{M}_1$ is then given in all subsets A of $\mathcal{X}$ by

$$\overline{E}_1(A) = \min\left\{\sum_{z \in A} \overline{m}_1(z), 1 - \sum_{z \in \mathcal{X} \setminus A} \underline{m}_1(z)\right\}, \tag{53}$$

This $\overline{E}_1$ is 2-*alternating*: $\overline{E}_1(A \cup B) + \overline{E}_1(A \cap B) \leq \overline{E}_1(A) + \overline{E}_1(B)$ for all subsets A and B of $\mathcal{X}$. This implies (see [33, Sect. 3.2.4] and [8, Thm. 8 and Cor. 17]) that for all $h \in \mathcal{L}(\mathcal{X})$, the upper expectation $\overline{E}_1(h)$ can be found by Choquet integration:

$$\overline{E}_1(h) = \min h + \int_{\min h}^{\max h} \overline{E}_1(\{z \in \mathcal{X} \colon h(z) \geq \alpha\})\, d\alpha, \tag{54}$$

where the integral is a Riemann integral. Similar considerations for the 2-alternating $\overline{E}_k(\cdot|x_k)$ lead to formulas for the upper transition operators $\overline{\mathrm{T}}_k$: For all x_k in $\mathcal{X}$,

$$\overline{\mathrm{T}}_k I_A(x_k) = \min\left\{\sum_{z \in A} \overline{q}_k(z|x_k), 1 - \sum_{z \in \mathcal{X} \setminus A} \underline{q}_k(z|x_k)\right\} \tag{55}$$

$$\overline{\mathrm{T}}_k h(x_k) = \min h + \int_{\min h}^{\max h} \overline{\mathrm{T}}_k I_{\{z \in \mathcal{X} \colon h(z) \geq \alpha\}}(x_k)\, d\alpha. \tag{56}$$

Using $\overline{E}_1$ and the $\overline{\mathrm{T}}_k$, all (conditional) expectations in the imprecise Markov chain can now be calculated, by applying Theorem 3.1 and Corollary 3.3.

Rather than using this backward recursion method, Škulj [31,32] used forward propagation, which, reformulated using our notations, amounts to the following. The marginal expectation $\overline{E}_2$ is calculated by $\overline{E}_2 = \overline{E}_1 \circ \overline{\mathrm{T}}_1$, $\overline{E}_3$ is calculated by $\overline{E}_3 = \overline{E}_2 \circ \overline{\mathrm{T}}_2$, and, more generally, $\overline{E}_{n+1} = \overline{E}_n \circ \overline{\mathrm{T}}_n$. Even though it appears quite natural, this approach has an important drawback, especially in the context of the probability interval approach described earlier. In order to calculate, say, $\overline{E}_3(h)$ we first need to find the upper expectation $\overline{E}_2$ and to calculate its value in the map $\overline{\mathrm{T}}_2 h$. However, $\overline{E}_2$, as the composition of two 2-alternating models $\overline{E}_1$ and $\overline{\mathrm{T}}_1$, is no longer necessarily 2-alternating and, therefore, its value in the map $\overline{\mathrm{T}}_2 h$ cannot generally be calculated from the values it assumes on events, using Choquet integration, as in Eqs (54) and (56). Indeed, Choquet integration will generally give too large a value for $\overline{E}_3(h)$

and will therefore lead to conservative approximations. These are the difficulties that Škulj was faced with in his work [31,32].

These difficulties can be circumvented by our backward recursion approach. Indeed, in order to find $\overline{E}_n(h)$, we begin by calculating $h_1 := h$ and $h_{k+1} := \overline{\mathrm{T}}_k h_k$, $k = 1, \dots, n-1$, using Eq. (56). Finally, $\overline{E}_n(h) = \overline{E}_1(h_n)$ is calculated using Eq. (54). Our calculations use Choquet integration but are tight, not conservative, approximations, because at all times the intervening local upper expectations are 2-alternating.

Example 6.4 (Close to a Cycle): Consider a three-state stationary imprecise Markov model with $\mathcal{X} = \{a, b, c\}$ and with marginal and transition probabilities given by probability intervals. It follows from Eqs (55) and (56) that the upper transition operator $\overline{\mathrm{T}}$ is fully determined by the lower and upper transition matrixes:

$$\underline{T} := \begin{bmatrix} \underline{q}(a|a) & \underline{q}(b|a) & \underline{q}(c|a) \\ \underline{q}(a|b) & \underline{q}(b|b) & \underline{q}(c|b) \\ \underline{q}(a|c) & \underline{q}(b|c) & \underline{q}(c|c) \end{bmatrix} = \frac{1}{200} \begin{bmatrix} 9 & 9 & 162 \\ 144 & 18 & 18 \\ 9 & 162 & 9 \end{bmatrix},$$

$$\overline{T} := \begin{bmatrix} \overline{q}(a|a) & \overline{q}(b|a) & \overline{q}(c|a) \\ \overline{q}(a|b) & \overline{q}(b|b) & \overline{q}(c|b) \\ \overline{q}(a|c) & \overline{q}(b|c) & \overline{q}(c|c) \end{bmatrix} = \frac{1}{200} \begin{bmatrix} 19 & 19 & 172 \\ 154 & 28 & 28 \\ 19 & 172 & 19 \end{bmatrix},$$

where the numerical values are particular to this example. We have depicted the credal sets $\mathcal{Q}(\cdot|a)$, $\mathcal{Q}(\cdot|b)$ and $\mathcal{Q}(\cdot|c)$ corresponding to this upper transition operator in Figure 7.

Similarly, the initial upper expectation $\overline{E}_1$ is completely determined by the row vectors $\underline{m}_1 := [\underline{m}_1(a)\ \ \underline{m}_1(b)\ \ \underline{m}_1(c)]$ and $\overline{m}_1 := [\overline{m}_1(a)\ \ \overline{m}_1(b)\ \ \overline{m}_1(c)]$. In Figure 8, we plot conservative approximations for the credal sets $\mathcal{M}_n$ corresponding to the upper expectation operators $\overline{E}_n$. Each approximation is based on the constraints that can be found by calculating $\underline{E}_1(\underline{\mathrm{T}}^{n-1} I_{\{x\}})$ and $\overline{E}_1(\overline{\mathrm{T}}^{n-1} I_{\{x\}})$ using the backward recursion method, for $x = a, b, c$. The $\mathcal{M}_n$ evolve clockwise through the simplex, which is not all that surprising, as the lower and upper transition matrixes are quite "close" to the precise *cyclic* transition matrix

$$T := \begin{bmatrix} q(a|a) & q(b|a) & q(c|a) \\ q(a|b) & q(b|b) & q(c|b) \\ q(a|c) & q(b|c) & q(c|c) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

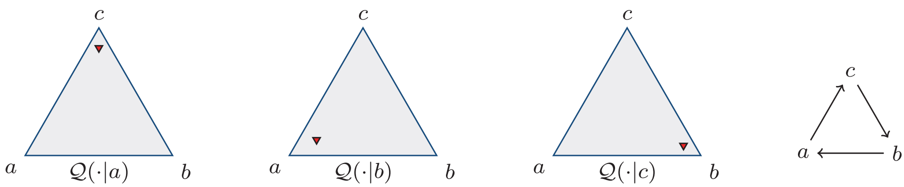

FIGURE 7. (Color online) The credal sets $\mathcal{Q}(\cdot|a)$, $\mathcal{Q}(\cdot|b)$, and $\mathcal{Q}(\cdot|c)$ in the simplex $\Sigma_{\{a,b,c\}}$, corresponding to the upper transition operator T in Example 6.4.

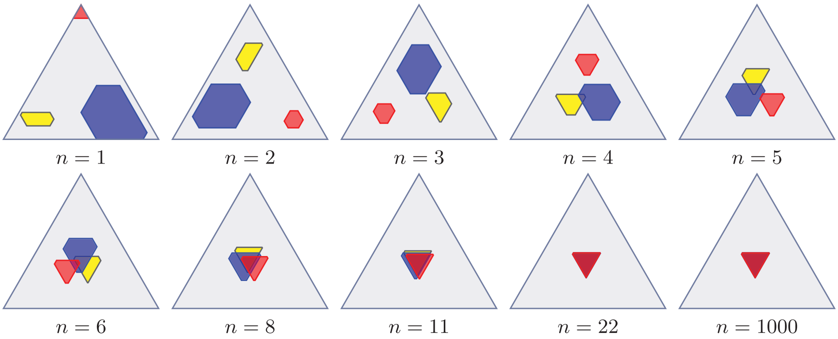

FIGURE 8. (Color online) Evolution in the simplex $\Sigma_{\{a,b,c\}}$ of the credal sets $\mathcal{M}_n$ for the near-cyclic transition operator from Example 6.4 for three different choices of the initial credal set $\mathcal{M}_1$.

as is also evident from Figure 7. After a while, the $\mathcal{M}_n$ converge to a limit that is independent of the initial credal set $\mathcal{M}_1$, as can be predicted from the regularity of the upper transition operator.

A biological application of imprecise Markov models can be found in Dhaenens's master's thesis [11]. He used the sensitivity analysis interpretation of imprecise Markov models to investigate the legitimacy of using PAM matrixes in amino acid and DNA sequence alignments. Roughly speaking, PAM (point accepted mutation) matrixes describe the chance that one amino acid mutates into another amino acid over a given evolutionary time span. However, the actual value of PAM matrix components are based on an estimation using an evolutionary model (i.e., amino acid substitutions are actually counted on the branches of a phylogenetic tree), hence the need to perform a sensitivity analysis. Dhaenens [11] observed in simulations that the imprecision due to the estimation did not blow up even after a large number of steps; he concluded that using PAM matrixes over large evolutionary timescales is still reasonable.

6.4. A *k*-out-of-*n*:F System With Uncertain Reliabilities

Reliability theory is one field in which Markov chains are used extensively. It concerns itself with questions of the type: What is the probability of failure of a system with n components? In the simplest case, where each component is either working or not working, answering this question would involve assessing the failure probabilities of the 2^n possible configurations of component states. However, as shown by Koutras [20], a great variety of reliability structures can be evaluated quite efficiently using their so-called embedded Markov chain. Among these are precisely those systems that fail as soon as any k out of the n components fail, also known as k-out-of-n:F systems.

For such systems, the embedded Markov chain is constructed as follows. Its state space $\mathcal{X}$ is given by $\{0, 1, 2, \ldots, k\}$, where each number represents the number of components that fail in the system. System failure is therefore represented by the event $\{k\}$ and a fully functioning is represented system by the event $\{0\}$. Koutras [20] showed that the failure probability (or unreliability) F_n and the reliability $R_n = 1 - F_n$ of a Markov chain embedded system are determined by the expectation form expression

$$F_n := E_{n+1}(I_{\{k\}}) = E_1(\mathrm{T}_1\mathrm{T}_2\cdots\mathrm{T}_nI_{\{k\}}), \tag{57}$$

where the initial distribution E_1 represents a system in perfect working condition, so $E_1(h) = h(0)$ for all real-valued maps h on $\mathcal{X}$. The transition matrix T_i corresponding to the transition operator T_i is fully determined by the reliability p_i of the ith component:

$$T_i = \begin{bmatrix} p_i & 1-p_i & 0 & \dots & 0 & 0 \\ 0 & p_i & 1-p_i & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & p_i & 1-p_i \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \tag{58}$$

where $(T_i)_{\ell,m} = \mathrm{T}_iI_{\{m\}}(\ell)$ and $\ell, m \in \{0, 1, \dots, k\}$.

Precise assessments of the individual reliabilities of the components p_i are often difficult to come by, as, for example, they might depend on climatological parameters, age, or even on the failure of other (external) components. However, experts might still be able to give conservative bounds on the individual reliabilities p_i. In this case, the embedded Markov chain becomes imprecise, but the corresponding bounds on the reliability and unreliability can still be computed by applying our sensitivity analysis formulas derived earlier:

$$\overline{F}_n = 1 - \underline{R}_n = E_1(\overline{\mathrm{T}}_1\overline{\mathrm{T}}_2\cdots\overline{\mathrm{T}}_nI_{\{k\}}) \quad \text{and} \quad \underline{F}_n = 1 - \overline{R}_n = E_1(\underline{\mathrm{T}}_1\underline{\mathrm{T}}_2\cdots\underline{\mathrm{T}}_nI_{\{k\}}). \tag{59}$$

When this embedded Markov chain is stationary (meaning that the uncertainty models for the reliability of all components are assumed to be the same), the failure probability bounds are simply computed by $\overline{F}_n = E_1(\overline{\mathrm{T}}^nI_{\{k\}})$ and $\underline{F}_n = E_1(\underline{\mathrm{T}}^nI_{\{k\}})$.

To give a very simple example, let us assume that an expert provides the same range $[\underline{r}, \overline{r}]$ for all component failure probabilities p_i, where $0 \leq \underline{r} \leq \overline{r} \leq 1$. This leads to a special case of the models considered in Section 6.3, and if we apply the formulas derived there, we get, after some manipulations, that

$$\overline{\mathrm{T}}h(\ell) = \begin{cases} \underline{r}h(\ell) + (1-\overline{r})h(\ell+1) \\ \quad + \; (\overline{r}-\underline{r})\max\{h(\ell), h(\ell+1)\} & \text{if } \ell = 0, 1, \dots, k-1 \\ h(k) & \text{if } \ell = k \end{cases} \tag{60}$$

for all real-valued maps h on $\mathcal{X}$. If h is nondecreasing in the sense that $h(0) \leq h(1) \leq \cdots \leq h(k-1) \leq h(k)$, then so is $\overline{\mathrm{T}}h$, and it therefore follows that

$$\overline{F}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{r} & 1-\underline{r} & 0 & \dots & 0 & 0 \\ 0 & \underline{r} & 1-\underline{r} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \underline{r} & 1-\underline{r} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}^n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{61}$$

$$= \sum_{\ell=k}^{n} \binom{n}{\ell} \underline{r}^{n-\ell}(1-\underline{r})^{\ell} = 1 - \sum_{\ell=0}^{k-1} \binom{n}{\ell} \underline{r}^{n-\ell}(1-\underline{r})^{\ell}, \tag{62}$$

and there is a completely similar expression for $\underline{F}_n$ where $\overline{r}$ is substituted for $\underline{r}$. See Figure 9 for a graphical illustration of these expressions.

If $0 < \underline{r} \leq \overline{r} \leq 1$, then this stationary imprecise Markov chain is regularly absorbing with regular top class $\{k\}$ (under $\rightarrow$), and $\underline{E}_\infty(h) = \overline{E}_\infty(h) = h(k)$ for all real-valued maps h on $\mathcal{X}$. Nevertheless, as soon as $\overline{r} = 1$, Hartfiel's product scrambling condition is no longer satisfied, as the identity matrix will then belong to all $\mathcal{T}_{\overline{\mathrm{T}}^n}$.

The chain ceases to be regularly absorbing if $\underline{r} = 0$ and $\overline{r} = 1$, and in that case, it is easy to see that $\overline{\mathrm{T}}^{k+n}h(m) = \max_{\ell=m}^{k} h(\ell)$ for all $n \geq 0$ and all real-valued maps h on $\mathcal{X}$; therefore, the limit upper expectation $\overline{E}_\infty$ will depend on the initial upper expectation $\overline{E}_1$. For the particular initial expectation E_1 we use in this example, we see that $\overline{E}_\infty(h) = \max h$.

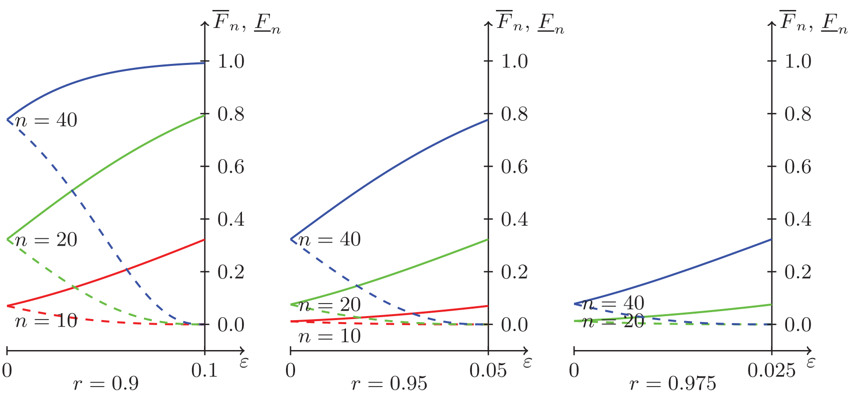

FIGURE 9. (Color online) Upper failure probability ($\overline{F}_n$, full line) and lower failure probability ($\underline{F}_n$, dashed line) for a 3-out-of-n:F system, for different numbers of components n as a function of the imprecision $\varepsilon := (\overline{r} - \underline{r})/2$ of the component reliability, for three different values of $r := (\overline{r} + \underline{r})/2$. As can be expected, the failure bounds widen with increasing imprecision, decrease with increasing reliability (characterised by r), and increase for a greater number of components n.

6.5. General Models

When the (conditional) upper expectation operators that define an imprecise Markov chain do not fall into any of the special cases we discussed and illustrated earlier, recourse must taken to more general calculation rules.

Let us consider the typical case of a credal set $\mathscr{P}$ that is specified by giving, for a finite number of real-valued maps f collected in the set $\mathcal{K} \subset \mathcal{L}(\mathcal{X})$, consistent upper bounds $U(f)$ on the expectations $E(f)$. Then the upper expectation for any map $h \in \mathcal{L}(\mathcal{X})$ can be found by solving the following linear program (see, e.g., [33, Sect. 3.1.3]):

$$\overline{E}_{\mathscr{P}}(h) = \min \left[\mu + \sum_{f \in \mathcal{K}} \lambda_f U(f) \right] \qquad \text{subject to} \quad h \leq \mu + \sum_{f \in \mathcal{K}} \lambda_f U(f) \qquad \textbf{(63)}$$
$$\text{where} \quad \lambda_f \geq 0 \text{ and } \mu \in \mathbb{R}.$$

As the number of upper expectations to compute, and thus the number of linear programs to solve, increases, it will eventually become profitable to take a second (dual) approach. Any credal set $\mathscr{P}$ specified by a finite number of constraints (bounds on expectations) is a convex polytope (i.e., has a finite set ext $\mathscr{P}$ of extreme points). Vertex enumeration algorithms such as the one by Avis, Bremner, and Seidel [1] can be used to obtain this set of extreme points from the given set of constraints. We can then use a practical version of Eq. (13) to find the corresponding upper expectations, namely (see [33, Sect. 3.1.3]):

$$\overline{E}_{\mathscr{P}}(h) := \max\left\{E_q(h) \colon q \in \operatorname{ext} \mathscr{P}\right\}. \qquad \textbf{(64)}$$

We can now consider imprecise Markov chains for which the local models, attached to the nonterminal situations in the tree, are of this type. The general backward recursion formulas we have given in Section 3 can then be used in combination with the formulas of the type (63) and (64) for the calculation of all conditional and joint upper and lower expectations in the tree.

7. CONCLUSIONS

To conclude, we (i) reflect on what type of convergence results could be obtained for imprecise Markov chains that are not regularly absorbing, (ii) pay attention to the important issue of interpretation of imprecise-probability models, and (iii) compare Hartfiel's approach [14] to our own regarding their practical applicability to deal with expectation problems.

It is a reasonably weak requirement for a stationary imprecise Markov chain with upper transition operator $\overline{\mathrm{T}}$ to be regularly absorbing, but we have seen that it is strong enough to guarantee that the upper expectation for the state at time n converges to a uniquely $\overline{\mathrm{T}}$-invariant upper expectation $\overline{E}_\infty$, regardless of the initial upper expectation $\overline{E}_1$.

Even when an imprecise Markov chain is not regularly absorbing, it is not so hard to see that its upper transition operator $\overline{\mathrm{T}}$ is still *nonexpansive* under the supremum norm given for every $h \in \mathcal{L}(\mathcal{X})$ by $\|h\|_\infty := \max|h|$, as

$$\|\overline{\mathrm{T}}g - \overline{\mathrm{T}}h\|_\infty \leq \|\overline{\mathrm{T}}(g-h)\|_\infty \leq \|g-h\|_\infty. \tag{65}$$

Moreover, the sequence $\|\overline{\mathrm{T}}^n h\|_\infty$ is bounded because $\|\overline{\mathrm{T}}^n h\|_\infty \leq \|h\|_\infty$. It then follows from nonlinear Perron–Frobenius theory [26,30] that the sequence $\overline{\mathrm{T}}^n h$ has a periodic limit cycle. More precisely, there is a $\xi_h \in \mathcal{L}(\mathcal{X})$ such that $\overline{\mathrm{T}}^{p_h}\xi_h = \xi_h$ (i.e., ξ_h is a *periodic point* of $\overline{\mathrm{T}}$ with (smallest) *period* p_h) and such that $\overline{\mathrm{T}}^{np_h}h \to \xi_h$ (pointwise) as $n \to \infty$. It would be a very interesting topic for further research to study the nature of the periods and periodic points of upper transition operators.

In our discussions, for instance in Section 3, we have consistently used the sensitivity analysis interpretation of imprecise-probability models such as upper expectations. Upper and lower expectations can also be given another, so-called *behavioral* interpretation, in terms of some subject's dispositions toward accepting risky transactions. This is, for instance, Walley's [33] preferred approach. The results we have derived here remain valid on that alternative interpretation, and the concatenation formulas (21) and (22) can then be shown to be special cases of the so-called *marginal extension* procedure [23], which provides the most conservative coherent (i.e., rational) inferences from the local predictive models $\overline{\mathrm{T}}_k$ to general lower and upper expectations. In another article [6], we give more details about how to approach a process theory using imprecise probabilities on a behavioral interpretation.

On a related matter, the imprecise Markov chains we are considering here can be seen as special *credal networks* [4,5,24]: the generalization of Bayesian networks to the case where the local models, associated with the nodes of the network, are credal sets. The corresponding "independence" notion that should then be used for the interpretation of the graphical structure of the network is Walley's *epistemic irrelevance* [33, Chap. 9]. Interestingly, Hartfiel's Markov set-chain approach corresponds to special credal nets for which the independence concept involved is a different one: that of *strong independence* [4]. Nevertheless, both approaches yield the same results if we restrict ourselves to calculating the marginal upper expectations for variables $X(n)$, as we have proved in Proposition 5.2. However, in any case, for the actual calculation of expectations, the set of transition matrixes approach suffers from a combinatorial explosion of computational complexity that can be avoided using our upper transition operator approach.

Acknowledgments

The authors wish to thank Damjan Škulj for inspirational discussion and two anonymous referees for helpful suggestions and pointers to the literature.

This article presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimisation), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

Notes

1. A *cut* V of a situation s is a collection of descendants v of s such that every path (from root to leaves) through s goes through exactly one v in V.

2. A *nonterminal* situation is a node of the tree that is not a leaf.

3. Arguments for the "expectation approach" to probability theory were given by Whittle [37]. This approach is also central in the work of Finetti [10]. For classical, precise probabilities, whether we use the language of probability measures or that of expectation operators seems to be a matter of personal preference, as the two approaches are formally equivalent. However, for the imprecise-probability models we introduce in Section 2, it was argued by Walley [33] that the (lower and upper) expectation language is mathematically superior and more expressive.

4. See Chapter 3 of Shafer [29] on causal reasoning in probability trees, which contains a number of propositions about calculating probabilities and expectations in probability trees. That such backward recursion is possible was arguably discovered by Christiaan Huygens in the middle of the 17th century. Shafer [29, Appendix A] discussed Huygens's treatment [16, Appendix VI] of a special case of the so-called *Problem of Points*, where Huygens drew what is probably the first recorded probability tree and solved the problem by backwards calculation of expectations in the tree.

5. The operators T_n are also called the *generators* of the Markov process; see Whittle [37].

6. The $\mathbb{T}^n$ can be seen as projection operators, since (with some abuse of notation) $\mathbb{T}_n \circ \mathbb{T}_n = \mathbb{T}_n$.

7. Also known as the Rule of Total Expectation, or the Rule of Total Probability, or the Conglomerative Property; see, for example, Whittle [37, Sect. 5.3] or Finetti [10].

8. In the literature we also find the term *stochastic matrix*; see Hartfiel [14], for instance.

9. This means that there is a $k > 0$ such that all elements of the kth power T^k of the transition matrixes T are (strictly) positive. Matrixes with this property are sometimes called *regular* as well, but this same name is also used for other matrix properties. Another name for this property is '*primitive*' [14].

10. These local transition probability masses themselves depend on the situation $x_{1:k}$ they are attached to, but the sets $\mathcal{Q}_k(\cdot|x_k)$ they are chosen from only depend on the last state x_k, as the Markov condition (16) tells us.

11. Usually, in Shafer's approach, Eq. (51) is only considered for (indicators of) events, and it then defines a so-called *plausibility function*, whose conjugate lower probability is a *belief function*. Equation (51) gives the pointwise greatest (most conservative) upper expectation that extends this plausibility function from events to real-valued maps.

References

1. Avis, D., Bremner, D., & Seidel, R. (1997). How good are convex hull algorithms? *Computational Geometry* 7(5–6): 265–301.
2. Campos, M.A., Dimuro, G.P., da Rocha Costa, A.C., & Kreinovich, V. (2003). Computing 2-step predictions for interval-valued finite stationary Markov chains. Technical report UTEP-CS-03-20a, University of Texas at El Paso.
3. de Campos, L.M., Huete, J.F., & Moral, S. (1994). Probability intervals: A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2: 167–196.
4. Cozman, F.G. (2000). Credal networks. *Artificial Intelligence* 120: 199–233.
5. Cozman, F.G. (2005). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning* 39(2–3): 167–184.
6. De Cooman, G. & Hermans, F. (2008). Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence* 172(11): 1400–1427.
7. De Cooman, G. & Miranda, E. (2007). Symmetry of models versus models of symmetry. In W.L. Harper & G.R. Wheeler (eds.), *Probability and inference: Essays in honor of Henry E. Kyburg, Jr.* London: King's College Publications, pp. 67–149.
8. De Cooman, G., Troffaes, M.C.M., & Miranda, E. (2008). n-Monotone exact functionals. *Journal of Mathematical Analysis and Applications* 347: 143–156.

9. de Finetti, B. (1970). *Teoria delle Probabilità*. Turin: Einaudi.
10. de Finetti, B. (1974–1975). *Theory of probability: A critical introductory treatment*. Chichester: Wiley. English translation of [9], two volumes.
11. Dhaenens, S. (2007). *Onderzoek van imprecieze Markov-modellen*. Master's thesis, Ghent University.
12. Harmanec, D. (2002). Generalizing Markov decision processes to imprecise probabilities. *Journal of Statistical Planning and Inference* 105: 199–213.
13. Hartfiel, D.J. (1991). Sequential limits in Markov set-chains. *Journal of Applied Probability* 28(4): 910–913.
14. Hartfiel, D.J. (1998). *Markov set-chains*. Lecture Notes in Mathematics No, 1695. Berlin: Springer.
15. Hartfiel, D.J. & Seneta, E. (1994). On the theory of Markov set-chains. *Advances in Applied Probability* 26: 947–964.
16. Huygens, C. (1656–1657). *Van Rekeningh in Spelen van Geluck*. Reprinted in Volume XIV of [17].
17. Huygens, C. (1888–1950). *Oeuvres complètes de Christiaan Huygens*. Den Haag: Martinus Nijhoff. Twenty-two volumes. Available in digitized form from the Bibliothèque nationale de France (http://gallica.bnf.fr).
18. Itoh, H. & Nakamura, K. (2007). Partially observable Markov decision processes with imprecise parameters. *Artificial Intelligence* 171: 453–490.
19. Kemeny, J.G. & Snell, J.L. (1976). *Finite markov chains*. Undergraduate Text in Mathematics. New York: Springer-Verlag.
20. Koutras, M.V. (1996). On a Markov chain approach for the study of reliability structures. *Journal of Applied Probability* 33(2): 357–367.
21. Kozine, I.O. & Utkin, L.V. (2002). Interval-valued finite Markov chains. *Reliable Computing* 8(2): 97–113.
22. Luenberger, D.G. (1979). *Introduction to dynamic systems. Theory, models & applications*. New York: Wiley.
23. Miranda, E. & de Cooman, G. (2007). Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning* 46(1): 188–225.
24. Moral, S. (2005). Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics and Artificial Intelligence* 45: 197–214.
25. Nilim, A. & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research* 53: 780–798.
26. Nussbaum, R.D., Scheutzow, M., & Verduyn Lunel, S.M. (1998). Periodic points of nonexpansive maps and nonlinear generalizations of the Perron–Frobenius theory. *Selecta Mathematica* 4: 141–181.
27. Satia, J.K. & Lave, R.E. (1973). Markovian decision processes with uncertain transition probabilities. *Operations Research* 21: 728–740.
28. Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
29. Shafer, G. (1996). *The art of causal conjecture*. Cambridge, MA: The MIT Press.
30. Sine, R. (1990). A nonlinear Perron–Frobenius theorem. *Proceedings of the American Mathematical Society* 109(2): 331–336.
31. Škulj, D. (2006). Finite discrete time Markov chains with interval probabilities. In J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegorzewski, & O. Hryniewicz (eds.), *Soft methods for integrated uncertainty modelling*. Berlin: Springer, pp. 299–306.
32. Škulj, D. (2007). Regular finite Markov chains with interval probabilities. In G. de Cooman, J. Vejnarová, & M. Zaffalon (eds.), *ISIPTA '07 — Proceedings of the fifth international symposium on imprecise probability: Theories and applications*. Prague: SIPTA, pp. 405–413.
33. Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman & Hall.
34. Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial Intelligence* 83(1): 1–58.
35. Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. I. Intervallwahrscheinlichkeit als umfassendes Konzept*. Heidelberg: Physica-Verlag.
36. White, C.C. & Eldeib, H.K. (1994). Markov decision-processes with imprecise transition-probabilities. *Operations Research* 42: 739–749.
37. Whittle, P. (2000). *Probability via expectation*. 4th ed. New York: Springer.

APPENDIX: PROOFS

In this Appendix, we have gathered proofs for the results in the article.

Before we go on, it will be useful to discuss and collect a number of properties of the upper transition operators associated with imprecise Markov chains. They follow immediately from the corresponding properties $\overline{E}1$–$\overline{E}7$ of upper expectations, so we omit the proof.

PROPOSITION A.1 (Properties of Upper Transition Operators): *Consider an imprecise Markov chain with a set of states $\mathcal{X}$ and upper transition operators $\overline{\mathrm{T}}_k$. Then for arbitrary h, h_1, h_2, and h_n in $\mathcal{L}(\mathcal{X})$, real $\lambda \geq 0$ and real μ; we have the following:*

- $\overline{\mathrm{T}}1$. $I_{\mathcal{X}} \min h \leq \overline{\mathrm{T}}_k h \leq I_{\mathcal{X}} \max h$ *(boundedness).*
- $\overline{\mathrm{T}}2$. $\overline{\mathrm{T}}_k(h_1 + h_2) \leq \overline{\mathrm{T}}_k h_1 + \overline{\mathrm{T}}_k h_2$ *(subadditivity).*
- $\overline{\mathrm{T}}3$. $\overline{\mathrm{T}}_k(\lambda h) = \lambda \overline{\mathrm{T}}_k h$ *(non-negative homogeneity).*
- $\overline{\mathrm{T}}4$. $\overline{\mathrm{T}}_k(h + \mu I_{\mathcal{X}}) = \overline{\mathrm{T}}_k h + \mu I_{\mathcal{X}}$ *(constant additivity).*
- $\overline{\mathrm{T}}5$. *If* $h_1 \leq h_2$, *then* $\overline{\mathrm{T}}_k h_1 \leq \overline{\mathrm{T}}_k h_2$ *(monotonicity).*
- $\overline{\mathrm{T}}6$. *If* $h_n \to h$ *pointwise, then* $\overline{\mathrm{T}}_k h_n \to \overline{\mathrm{T}}_k h$ *pointwise (continuity).*
- $\overline{\mathrm{T}}7$. $\overline{\mathrm{T}}_k h \geq -\overline{\mathrm{T}}_k(-h) = \underline{\mathrm{T}}_k h$ *(upper–lower consistency).*

Consider any operator $\overline{\mathrm{T}}\colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ that satisfies properties $\overline{\mathrm{T}}1$–$\overline{\mathrm{T}}3$. Then for each $x \in \mathcal{X}$, the real functional $\overline{E}(\cdot|x)$ defined on $\mathcal{L}(\mathcal{X})$ by $\overline{E}(h|x) = \overline{\mathrm{T}}h(x)$ is an upper expectation, because it satisfies properties $\overline{E}1$–$\overline{E}3$. This means that we can consider $\overline{\mathrm{T}}$ as an upper transition operator associated with some imprecise Markov chain. It therefore make sense to call any operator $\overline{\mathrm{T}}$ that satisfies properties $\overline{\mathrm{T}}1$–$\overline{\mathrm{T}}3$ an *upper transition operator*. Clearly, if $\overline{\mathrm{T}}_1, \dots, \overline{\mathrm{T}}_n$ are upper transition operators, then so is their composition $\overline{\mathrm{T}}_1 \cdots \overline{\mathrm{T}}_n$.

We are now ready to proceed with the proofs of all results in the body of the article.

PROOF OF THEOREM 3.1: We first prove by induction that the left-hand sides are dominated by the right-hand sides in Eqs (21) and (22). To get the induction process started, we observe that Eq. (21) holds trivially for $n = N - 1$. Next, we prove that if the desired inequality in Eq. (21) holds for $n = k + 1$, it also holds for $n = k$, where k is any element in $\{1, 2, \dots, N - 2\}$. Let us fix $x_{1:k} \in \mathcal{X}^k$; then we have to prove that

$$\overline{E}(f|x_{1:k}) \leq \overline{\mathbb{T}}_k \overline{\mathbb{T}}_{k+1} \cdots \overline{\mathbb{T}}_{N-1} f(x_{1:k}), \tag{A.1}$$

where we can use that, in particular, for all $x_{k+1} \in \mathcal{X}$,

$$\overline{E}(f|x_{1:k}, x_{k+1}) \leq \overline{\mathbb{T}}_{k+1} \overline{\mathbb{T}}_{k+2} \cdots \overline{\mathbb{T}}_{N-1} f(x_{1:k}, x_{k+1}). \tag{A.2}$$

We have fixed $x_{1:k}$, so we can regard $\overline{E}(f|x_{1:k}, \cdot)$ as a real-valued map on $\mathcal{X}$, depending only on the state $X(k + 1)$ at time $k + 1$. We denote this map by h_{k+1}.

Now, consider any compatible probability tree. In particular, let $q(\cdot|x_{1:k}) \in \mathcal{Q}_k(\cdot|x_k)$ be the corresponding local probability mass function for the uncertainty about the state $X(k + 1)$

in the situation $x_{1:k}$ we are considering. It follows from the Law of Iterated Expectations that in this probability tree,

$$E(f|x_{1:k}) = E(E(f|x_{1:k},\cdot)|x_{1:k}), \quad \textbf{(A.3)}$$

and since $E(f|x_{1:k},\cdot) \leq \overline{E}(f|x_{1:k},\cdot) = h_{k+1}$, by definition of the upper expectations in the tree, we may derive from the monotonicity of expectation operators that $E(f|x_{1:k}) \leq E(h_{k+1}|x_{1:k})$. Now, h_{k+1} is a function of $X(k+1)$ only, so its conditional expectation $E(h_{k+1}|x_{1:k})$ in situation $x_{1:k}$ can be calculated using the local conditional model $q(\cdot|x_{1:k})$ for $X(k+1)$; that is,

$$E(h_{k+1}|x_{1:k}) = \sum_{x_{k+1}\in\mathcal{X}} h_{k+1}(x_{k+1})q(x_{k+1}|x_{1:k}) \leq \overline{E}_k(h_{k+1}|x_k), \quad \textbf{(A.4)}$$

where the inequality follows from Eq. (15). Hence, $E(f|x_{1:k}) \leq \overline{E}_k(h_{k+1}|x_k)$ and, therefore,

$$\begin{aligned}\overline{E}(f|x_{1:k}) &\leq \overline{E}_k(h_{k+1}|x_k) = \overline{\mathrm{T}}_k h_{k+1}(x_k)\\ &\leq \overline{\mathrm{T}}_k\left(\overline{\mathbb{T}}_{k+1}\overline{\mathbb{T}}_{k+2}\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:k},\cdot)\right)(x_k) = \overline{\mathbb{T}}_k\overline{\mathbb{T}}_{k+1}\overline{\mathbb{T}}_{k+2}\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:k}), \quad \textbf{(A.5)}\end{aligned}$$

where the first inequality follows from the definition of the upper expectations in the tree, the first equality follows from Eq. (19), the second inequality follows from Eq. (A.2) and the monotonicity ($\overline{\mathrm{T}}5$) of upper transition operators, and the second equality follows from Eq. (20).

In a completely similar way, but now using the model $\mathcal{M}_1$ rather than the model $\mathcal{Q}_k(\cdot|x_k)$, we can prove that the desired inequalities hold for $n = 0$, given that they hold for $n = 1$. So now we know that the left-hand sides are dominated by the right-hand sides in Eqs (21) and (22).

It remains to prove the converse inequalities. Fix any path in the tree. We denote the successive situations on this path by $\square, x_{1:1}, x_{1:2}, \ldots, x_{1:N-1}, x_{1:N}$. First, consider the situation $x_{1:N-1}$ and the partial map $h_N := f(x_{1:N-1},\cdot)$, then we know, because the credal set $\mathcal{Q}_{N-1}(\cdot|x_{N-1})$ is convex and closed, that there is some probability mass function in $\mathcal{Q}_{N-1}(\cdot|x_{N-1})$, which we denote by $\hat{q}(\cdot|x_{1:N-1})$, such that

$$\begin{aligned}\sum_{x_N\in\mathcal{X}} h_N(x_N)\hat{q}(x_N|x_{1:N-1}) = \overline{E}_{N-1}(h_N|x_{N-1}) &= \overline{\mathrm{T}}_{N-1}f(x_{1:N-1},\cdot)(x_{N-1})\\ &= \overline{\mathbb{T}}_{N-1}f(x_{1:N-1}) \quad \textbf{(A.6)}\end{aligned}$$

and, therefore,

$$\overline{\mathbb{T}}_{N-1}f(x_{1:N-1}) = \sum_{x_N\in\mathcal{X}} f(x_{1:N-1},x_N)\hat{q}(x_N|x_{1:N-1}). \quad \textbf{(A.7)}$$

Next, consider the situation $x_{1:N-2}$ and the partial map $h_{N-1} := \overline{\mathbb{T}}_{N-1}f(x_{1:N-2},\cdot)$. Again, we know since $\mathcal{Q}_{N-2}(\cdot|x_{N-2})$ is convex and closed that there is some probability mass function in $\mathcal{Q}_{N-2}(\cdot|x_{N-2})$, which we denote by $\hat{q}(\cdot|x_{1:N-2})$, such that

$$\begin{aligned}\sum_{x_{N-1}\in\mathcal{X}} h_{N-1}(x_{N-1})\hat{q}(x_{N-1}|x_{1:N-2}) &= \overline{E}_{N-2}(h_{N-1}|x_{N-2})\\ &= \overline{\mathrm{T}}_{N-2}\left(\overline{\mathbb{T}}_{N-1}f(x_{1:N-2},\cdot)\right)(x_{N-2})\\ &= \overline{\mathbb{T}}_{N-2}\overline{\mathbb{T}}_{N-1}f(x_{1:N-2}) \quad \textbf{(A.8)}\end{aligned}$$

and, therefore,

$$\sum_{x_{N-1}\in\mathcal{X}} \overline{\mathbb{T}}_{N-1}f(x_{1:N-2},x_{N-1})\hat{q}(x_{N-1}|x_{1:N-2}) = \overline{\mathbb{T}}_{N-2}\overline{\mathbb{T}}_{N-1}f(x_{1:N-2}). \tag{A.9}$$

If we combine Eqs (A.7) and (A.9), we find that

$$\sum_{x_{N-1:N}\in\mathcal{X}^2} f(x_{1:N-2},x_{N-1:N})\hat{q}(x_{N-1}|x_{1:N-2})\hat{q}(x_N|x_{1:N-1}) = \overline{\mathbb{T}}_{N-2}\overline{\mathbb{T}}_{N-1}f(x_{1:N-2}). \tag{A.10}$$

We can obviously continue in this manner until we reach the root of the tree. We have then effectively constructed a compatible probability tree for which the associated conditional and joint expectation operators satisfy, for all situations ($n = 1,\ldots,N-1$),

$$\begin{aligned}\overline{E}(f|x_{1:n}) \geq \hat{E}(f|x_{1:n}) &:= \sum_{x_{n+1:N}\in\mathcal{X}^{N-n}} f(x_{1:n},x_{n+1:N})\prod_{k=n}^{N-1}\hat{q}(x_{k+1}|x_{1:k}) \\ &= \overline{\mathbb{T}}_n\overline{\mathbb{T}}_{n+1}\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:n}),\end{aligned} \tag{A.11}$$

$$\overline{E}(f) \geq \hat{E}(f) := \sum_{x_{1:N}\in\mathcal{X}^N} f(x_{1:N})\hat{m}_1(x_1)\prod_{k=1}^{N-1}\hat{q}(x_{k+1}|x_{1:k}) = \overline{E}_1(\overline{\mathbb{T}}_1\overline{\mathbb{T}}_2\cdots\overline{\mathbb{T}}_{N-1}f). \tag{A.12}$$

This tells us that the converse inequalities in Eqs (21) and (22) hold as well. ■

Proof of Proposition 3.2: We use Eq. (21). It is clear from the definition (20) of the $\overline{\mathbb{T}}_k$ that if f is $\{n,n+1,\ldots,N\}$-measurable, then $\overline{\mathbb{T}}_{N-1}f$ is $\{n,n+1,\ldots,N-1\}$-measurable and then $\overline{\mathbb{T}}_{N-2}\overline{\mathbb{T}}_{N-1}f$ is also $\{n,n+1,\ldots,N-2\}$-measurable; so by continuing the induction, we find $\overline{\mathbb{T}}_{n+1}\cdots\overline{\mathbb{T}}_{N-1}f$ is $\{n,n+1\}$-measurable and, finally, $\overline{\mathbb{T}}_n\cdots\overline{\mathbb{T}}_{N-1}f$ is $\{n\}$-measurable. ■

Proof of Corollary 3.3: We use Eqs (21) and (22) with f defined as follows: $f(x_{1:N}) := h(x_n)$ for all $x_{1:N}\in\mathcal{X}^N$. Then, also using property $\overline{\mathrm{T}}_3$, the nonnegative homogeneity of upper transition operators, we find after subsequently applying $\overline{\mathbb{T}}_{N-1},\ldots,\overline{\mathbb{T}}_\ell$ that

$$\begin{aligned}\overline{\mathbb{T}}_{N-1}f(x_{1:N-1}) &= \overline{\mathrm{T}}_{N-1}(h(x_n)I_{\mathcal{X}})(x_{N-1}) = h(x_n),\\ &\vdots\\ \overline{\mathbb{T}}_n\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:n}) &= \overline{\mathrm{T}}_n(h(x_n)I_{\mathcal{X}})(x_n) = h(x_n),\\ \overline{\mathbb{T}}_{n-1}\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:n-1}) &= \overline{\mathrm{T}}_{n-1}h(x_{n-1}),\\ \overline{\mathbb{T}}_{n-2}\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:n-2}) &= \overline{\mathrm{T}}_{n-2}\overline{\mathrm{T}}_{n-1}h(x_{n-2}),\\ &\vdots\\ \overline{\mathbb{T}}_\ell\cdots\overline{\mathbb{T}}_{N-1}f(x_{1:\ell}) &= \overline{\mathrm{T}}_\ell\overline{\mathrm{T}}_{\ell+1}\cdots\overline{\mathrm{T}}_{n-1}h(x_\ell)\end{aligned} \tag{A.13}$$

and, therefore, $\overline{\mathbb{T}}_\ell\ldots\overline{\mathbb{T}}_{N-1}f(x_{1:\ell-1},\cdot) = \overline{\mathrm{T}}_\ell\overline{\mathrm{T}}_{\ell+1}\ldots\overline{\mathrm{T}}_{n-1}h$. Applying Proposition 3.2 then leads to the first desired equality. If, for $\ell = 1$, we now also apply the upper expectation $\overline{E}_1$ to both sides of this equality, the proof is complete. ■

PROOF OF PROPOSITION 3.4: As an example, we prove Eq. (24), by applying Eq. (21) with its parameters chosen as $f = I_{\{x_{n+1:m}\}}$ and $N = m$. We then see that for any $z_{1:m-1} \in \mathcal{X}^{m-1}$,

$$\begin{aligned}\overline{\mathbb{T}}_{m-1} I_{\{x_{n+1:m}\}}(z_{1:m-1}) &= \overline{\mathrm{T}}_{m-1}\left(I_{\{x_{n+1:m-1}\}}(z_{n+1:m-1}) I_{\{x_m\}}\right)(z_{m-1}) \\ &= I_{\{x_{n+1:m-1}\}}(z_{n+1:m-1}) \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(z_{m-1}) \\ &= I_{\{x_{n+1:m-1}\}}(z_{n+1:m-1}) \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1}),\end{aligned} \tag{A.14}$$

where we have used the nonnegative homogeneity $\overline{\mathrm{T}}3$ of upper transition operators. Therefore $\overline{\mathbb{T}}_{m-1} I_{\{x_{n+1:m}\}} = I_{\{x_{n+1:m-1}\}} \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1})$. Consequently, for any $z_{1:m-2} \in \mathcal{X}^{m-2}$,

$$\begin{aligned}\overline{\mathbb{T}}_{m-2}\overline{\mathbb{T}}_{m-1} I_{\{x_{n+1:m}\}}(z_{1:m-2}) &= \overline{\mathrm{T}}_{m-2}\left(\overline{\mathbb{T}}_{m-1} I_{\{x_{n+1:m}\}}(z_{1:m-2})\right)(z_{m-2}) \\ &= \overline{\mathrm{T}}_{m-2}\left(I_{\{x_{n+1:m-2}\}}(z_{n+1:m-2}) I_{\{x_{m-1}\}} \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1})\right)(z_{m-2}) \\ &= I_{\{x_{n+1:m-2}\}}(z_{n+1:m-2}) \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1}) \overline{\mathrm{T}}_{m-2} I_{\{x_{m-1}\}}(z_{m-2}) \\ &= I_{\{x_{n+1:m-2}\}}(z_{n+1:m-2}) \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1}) \overline{\mathrm{T}}_{m-2} I_{\{x_{m-1}\}}(x_{m-2}),\end{aligned} \tag{A.15}$$

again using property $\overline{\mathrm{T}}3$, and, therefore,

$$\overline{\mathbb{T}}_{m-2}\overline{\mathbb{T}}_{m-1} I_{\{x_{n+1:m}\}} = I_{\{x_{n+1:m-2}\}} \overline{\mathrm{T}}_{m-1} I_{\{x_m\}}(x_{m-1}) \overline{\mathrm{T}}_{m-2} I_{\{x_{m-1}\}}(x_{m-2}). \tag{A.16}$$

Continuing in this fashion eventually leads to Eq. (24). ■

PROOF OF PROPOSITION 4.3: Suppose $\mathcal{R}_{\rightsquigarrow} \neq \emptyset$. Consider any maximal state y [there always is at least one, because $\mathcal{X}$ is finite] and any $x \in \mathcal{R}_{\rightsquigarrow}$; then it is clear from the definition of $\mathcal{R}_{\rightsquigarrow}$ that $y \rightsquigarrow x$. Since y is maximal, it follows that also $x \rightsquigarrow y$ and, therefore, $x \leftrightsquigarrow y$. We conclude that $\mathcal{R}_{\rightsquigarrow}$ is included in all maximal communication classes. This means that there is only one such maximal class, and $\mathcal{R}_{\rightsquigarrow}$ is included in this top class. To show that $\mathcal{R}_{\rightsquigarrow}$ is equal to this top class, consider any maximal element y and any $x \in \mathcal{R}_{\rightsquigarrow}$. Then we know that there is some $n \in \mathbb{N}$ such that for all $k \geq n$ and all $z \in \mathcal{X}$, $z \overset{k}{\rightsquigarrow} x$. However, we have seen earlier that $x \leftrightsquigarrow y$, so there is some $\ell \geq 0$ such that $x \overset{\ell}{\rightsquigarrow} y$ and, therefore, $z \overset{k+\ell}{\rightsquigarrow} y$ for all $z \in \mathcal{X}$. This implies that $y \in \mathcal{R}_{\rightsquigarrow}$, so $\mathcal{R}_{\rightsquigarrow}$ is indeed the top class. We show that it is regular. For each x in $\mathcal{R}_{\rightsquigarrow}$ there is an $n_x \in \mathbb{N}$ such that $y \overset{k}{\rightsquigarrow} x$ for all $k \geq n_x$ and all $y \in \mathcal{X}$. If we define $n := \max\{n_x : x \in \mathcal{R}_{\rightsquigarrow}\}$, then we see that $x \overset{k}{\rightsquigarrow} y$ for all $k \geq n$ and all $x, y \in \mathcal{R}_{\rightsquigarrow}$, so $\mathcal{R}_{\rightsquigarrow}$ is regular by Proposition 4.2 and, therefore, $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ is top class regular.

Conversely, assume that $\cdot \overset{\cdot}{\rightsquigarrow} \cdot$ is top class regular. Consider any state x in the top class and any $y \in \mathcal{X}$. Then there is some $\ell_y \geq 0$ such that $y \overset{\ell_y}{\rightsquigarrow} x$, and it follows from Proposition 4.2 that there is some $n \in \mathbb{N}$ such that $x \overset{k}{\rightsquigarrow} x$ and, therefore, $y \overset{\ell_y+k}{\rightsquigarrow} x$ for all $k \geq n$. So if we let $m := n + \max\{\ell_y : y \in \mathcal{X}\}$, then we see that $y \overset{k}{\rightsquigarrow} x$ for all $k \geq m$ and all $y \in \mathcal{X}$ and, therefore, $x \in \mathcal{R}_{\rightsquigarrow}$, from which $\mathcal{R}_{\rightsquigarrow} \neq \emptyset$. ■

PROOF OF PROPOSITION 4.4: Fix x, y, and z in $\mathcal{X}$. Since $\overline{P}^m_{uy} = \overline{\mathrm{T}}^m I_{\{y\}}(u) \geq 0$ for all $u \in \mathcal{X}$, we have that

$$\overline{\mathrm{T}}^m I_{\{y\}} = \sum_{u \in \mathcal{X}} \overline{\mathrm{T}}^m I_{\{y\}}(u) I_{\{u\}} \geq \overline{\mathrm{T}}^m I_{\{y\}}(z) I_{\{z\}}. \tag{A.17}$$

If we now apply the upper transition operator $\overline{\mathrm{T}}$ n times to both sides of this inequality and repeatedly invoke its monotonicity ($\overline{\mathrm{T}}_5$) and nonnegative homogeneity ($\overline{\mathrm{T}}_3$), we find that $\overline{\mathrm{T}}^{n+m} I_{\{y\}} \geq \overline{\mathrm{T}}^m I_{\{y\}}(z) \overline{\mathrm{T}}^n I_{\{z\}}$ and, hence, indeed $\overline{\mathrm{T}}^{n+m} I_{\{y\}}(x) \geq \overline{\mathrm{T}}^n I_{\{z\}}(x) \overline{\mathrm{T}}^m I_{\{y\}}(z)$. ∎

PROOF OF PROPOSITION 4.5: Fix x in $\mathcal{X}$. Boundedness ($\overline{\mathrm{T}}_1$) and subadditivity ($\overline{\mathrm{T}}_2$) guarantee that $0 < 1 \leq \overline{\mathrm{T}}^n I_{\mathcal{X}}(x) \leq \sum_{y \in \mathcal{X}} \overline{\mathrm{T}}^n I_{\{y\}}(x)$. So there must be some $y \in \mathcal{X}$ for which $\overline{P}^n_{xy} = \overline{\mathrm{T}}^n I_{\{y\}}(x) > 0$. ∎

The following lemma provides a characterization for top class regularity (under $\rightarrow$) that is somewhat simpler than the one implicit in Proposition 4.3.

LEMMA A.2: *A stationary imprecise Markov chain is top class regular (under $\rightarrow$) if and only if*

$$\mathcal{R}_{\rightarrow} = \{x \in \mathcal{X} \colon (\exists n \in \mathbb{N})(\forall y \in \mathcal{X}) y \overset{n}{\rightarrow} x\} \neq \emptyset. \tag{A.18}$$

PROOF: Let $\mathcal{R}'_{\rightarrow} := \{x \in \mathcal{X} \colon (\exists n \in \mathbb{N})(\forall y \in \mathcal{X}) y \overset{n}{\rightarrow} x\}$; then by Proposition 4.3 it suffices to prove that $\mathcal{R}_{\rightarrow} = \mathcal{R}'_{\rightarrow}$. It is clear that $\mathcal{R}_{\rightarrow} \subseteq \mathcal{R}'_{\rightarrow}$, so we concentrate on the converse inequality. Consider any $x \in \mathcal{X}$ and $n \in \mathbb{N}$ such that $y \overset{n}{\rightarrow} x$ for all $y \in \mathcal{X}$. Then it suffices to prove that also $y \overset{n+1}{\rightarrow} x$ for all $y \in \mathcal{X}$. Fix y; then there is some $z \in \mathcal{X}$ such that $\overline{P}^1_{yz} > 0$, by Proposition 4.5. However, since we know that for this z, $\overline{P}^n_{zx} > 0$, also we infer from Proposition 4.4 that indeed $\overline{P}^{n+1}_{yx} \geq \overline{P}^1_{yz} \overline{P}^n_{zx} > 0$. ∎

Before we come to the upper expectation form of the Perron–Frobenius theorem (Theorem 5.1), we first prove the following lemmas.

LEMMA A.3: *Let $\overline{\mathrm{T}}$ be an upper transition operator associated with some stationary imprecise Markov chain, meaning that it satisfies properties $\overline{\mathrm{T}}_1$–$\overline{\mathrm{T}}_7$. Consider any $h \in \mathcal{L}(\mathcal{X})$. Then the real sequence* $\min \overline{\mathrm{T}}^n h$, *$n \in \mathbb{N}$, is nondecreasing and converges to some limit $l(h) \in \mathbb{R}$. Similarly, the real sequence* $\max \overline{\mathrm{T}}^n h$, *$n \in \mathbb{N}$, is nonincreasing and converges to some limit $L(h) \in \mathbb{R}$. Of course,* $\min h \leq l(h) \leq L(h) \leq \max h$.

PROOF: Fix h in $\mathcal{L}(\mathcal{X})$ and consider any n in $\mathbb{N}_0$. From $I_{\mathcal{X}} \min \overline{\mathrm{T}}^n h \leq \overline{\mathrm{T}}^n h \leq I_{\mathcal{X}} \max \overline{\mathrm{T}}^n h$ (by property $\overline{\mathrm{T}}_1$) we deduce using property $\overline{\mathrm{T}}_5$ that $\overline{\mathrm{T}}(I_{\mathcal{X}} \min \overline{\mathrm{T}}^n h) \leq \overline{\mathrm{T}}^{n+1} h \leq \overline{\mathrm{T}}(I_{\mathcal{X}} \max \overline{\mathrm{T}}^n h)$ and, therefore, using properties $\overline{\mathrm{T}}_3$ and $\overline{\mathrm{T}}_4$, that $I_{\mathcal{X}} \min \overline{\mathrm{T}}^n h \leq \overline{\mathrm{T}}^{n+1} h \leq I_{\mathcal{X}} \max \overline{\mathrm{T}}^n h$. Consequently,

$$\min h \leq \min \overline{\mathrm{T}}^n h \leq \min \overline{\mathrm{T}}^{n+1} h \leq \max \overline{\mathrm{T}}^{n+1} h \leq \max \overline{\mathrm{T}}^n h \leq \max h. \tag{A.19}$$

This tells us that the real sequence $\max \overline{\mathrm{T}}^n h$ is nonincreasing and bounded below (by $\min h$). It therefore converges to some real number $L(h)$. Similarly, the real sequence $\min \overline{\mathrm{T}}^n h$ is nondecreasing and bounded above (by $\max h$) and therefore converges to some real number $l(h)$. That $\min h \leq l(h) \leq L(h) \leq \max h$ follows from the inequalities in Eq. (A.19) by taking the limit $n \rightarrow \infty$. ∎

LEMMA A.4: *Let $\overline{\mathrm{T}}$ be an upper transition operator associated with some stationary imprecise Markov chain, meaning that it satisfies properties $\overline{\mathrm{T}}_1$–$\overline{\mathrm{T}}_7$. Consider any $h \in \mathcal{L}(\mathcal{X})$. Then there is some x_o in $\mathcal{X}$ such that for all $n \in \mathbb{N}$, there is some $k_n > n$ for which $L(h) \leq \overline{\mathrm{T}}^{k_n} h(x_o)$. Moreover,* $\lim_{n\to\infty} \overline{\mathrm{T}}^{k_n} h(x_o) = \limsup_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) = L(h)$.

PROOF: Suppose, *ex absurdo*, that for any $x \in \mathcal{X}$ there is some $n_x \in \mathbb{N}$ such that for all $k > n_x$, $\overline{\mathrm{T}}^k h(x) < L(h)$. Since $\mathcal{X}$ is finite, this implies that there is some $n := \max\{n_x : x \in \mathcal{X}\}$ such that for all $k > n$, $\max \overline{\mathrm{T}}^k h < L(h)$. This contradicts the conclusion $\max \overline{\mathrm{T}}^n h \searrow L(h)$ obtained in Lemma A.3.

Next, we show that $\lim_{n\to\infty} \overline{\mathrm{T}}^{k_n} h(x_o) = L(h)$. For all $n \in \mathbb{N}$, $L(h) \leq \overline{\mathrm{T}}^{k_n} h(x_o) \leq \max \overline{\mathrm{T}}^{k_n} h$, and since the subsequence $\max \overline{\mathrm{T}}^{k_n} h$ converges to the same limit $L(h)$ as the convergent sequence $\max \overline{\mathrm{T}}^n h$, we see that the sequence $\overline{\mathrm{T}}^{k_n} h(x_o)$ converges to $L(h)$ as well.

To conclude, we show that $\limsup_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) = L(h)$. Since the limit superior of a sequence is the supremum of the limits of all its convergent subsequences and since, moreover, we have just proved that $\lim_{n\to\infty} \overline{\mathrm{T}}^{k_n} h(x_o) = L(h)$, we infer that $\limsup_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) \geq L(h)$. For the converse inequality, starting from $\overline{\mathrm{T}}^n h(x_o) \leq \max \overline{\mathrm{T}}^n h$ and taking the limit superior on both sides of the inequality yields $\limsup_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) \leq \limsup_{n\to\infty} \max \overline{\mathrm{T}}^n h = L(h)$, where the equality follows from Lemma A.3. ∎

LEMMA A.5: *Let $\overline{\mathrm{T}}$ be an upper transition operator associated with some stationary imprecise Markov chain, meaning that it satisfies properties $\overline{\mathrm{T}}_1$–$\overline{\mathrm{T}}_7$. Consider any $h \in \mathcal{L}(\mathcal{X})$. If the imprecise Markov chain is regularly absorbing, then $l(h) = L(h)$.*

PROOF: Since the imprecise Markov chain is in particular top class regular (under $\rightarrow$), we have by Proposition 4.3 that $\mathcal{R}_{\rightarrow} \neq \emptyset$. Consider any $x \in \mathcal{R}_{\rightarrow}$; then we first prove that $\lim_{n\to\infty} \overline{\mathrm{T}}^n h(x) = l(h)$. We know from the definition of $\mathcal{R}_{\rightarrow}$ that there is some $n_x \in \mathbb{N}$ such that $\min \overline{\mathrm{T}}^{n_x} I_{\{x\}} > 0$. Additionally, for any $n \geq 0$,

$$0 \leq \left[\overline{\mathrm{T}}^n h(x) - \min \overline{\mathrm{T}}^n h\right] I_{\{x\}} \leq \overline{\mathrm{T}}^n h - \min \overline{\mathrm{T}}^n h, \tag{A.20}$$

and if we apply $\overline{\mathrm{T}}$ n_x times to all sides of these inequalities, we get

$$0 \leq \left[\overline{\mathrm{T}}^n h(x) - \min \overline{\mathrm{T}}^n h\right] \overline{\mathrm{T}}^{n_x} I_{\{x\}} \leq \overline{\mathrm{T}}^{n+n_x} h - \min \overline{\mathrm{T}}^n h \tag{A.21}$$

after repeated use of properties $\overline{\mathrm{T}}_5$, $\overline{\mathrm{T}}_4$, and $\overline{\mathrm{T}}_3$. Taking the minimum of all sides of these inequalities leads to

$$0 \leq \left[\overline{\mathrm{T}}^n h(x) - \min \overline{\mathrm{T}}^n h\right] \min \overline{\mathrm{T}}^{n_x} I_{\{x\}} \leq \min \overline{\mathrm{T}}^{n+n_x} h - \min \overline{\mathrm{T}}^n h. \tag{A.22}$$

If we now let $n \to \infty$, we see that since the term on the right converges to zero [see Lemma A.3], so must the middle term. Since $\min \overline{\mathrm{T}}^{n_x} I_{\{x\}} > 0$, this implies that $\overline{\mathrm{T}}^n h(x) - \min \overline{\mathrm{T}}^n h$ converges to zero, from which indeed $\lim_{n\to\infty} \overline{\mathrm{T}}^n h(x) = \lim_{n\to\infty} \min \overline{\mathrm{T}}^n h = l(h)$.

As a next step, we infer from Lemma A.4 that there is some x_o in $\mathcal{X}$ and some strictly increasing sequence k_n of natural numbers, such that $L(h) \leq \overline{\mathrm{T}}^{k_n} h(x_o)$ for all $n \in \mathbb{N}$ and, moreover, $\limsup_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) = L(h)$.

There are now two possibilities. The first is that $x_o \in \mathcal{R}_{\rightarrow}$. Then it follows from the above discussion that $\lim_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) = l(h)$. However, since we also have that $\lim_{n\to\infty} \overline{\mathrm{T}}^n h(x_o) =$

$\lim_{n\to\infty} \overline{\mathrm{T}}^{k_n} h(x_o) = L(h)$, where the last equality follows from Lemma A.4, we infer that in this case indeed $l(h) = L(h)$.

The second possibility is that $x_o \notin \mathcal{R}_\rightarrow$, but then it follows from the assumption that there is some $n_o \in \mathbb{N}$ such that $\underline{\mathrm{T}}^{n_o} I_{\mathcal{R}_\rightarrow}(x_o) > 0$. We have for all $n \in \mathbb{N}$ that

$$0 \leq \left[\max \overline{\mathrm{T}}^n h - \max_{y \in \mathcal{R}_\rightarrow} \overline{\mathrm{T}}^n h(y)\right] I_{\mathcal{R}_\rightarrow} \leq \max \overline{\mathrm{T}}^n h - \overline{\mathrm{T}}^n h, \tag{A.23}$$

and if we apply $\underline{\mathrm{T}}$ n_o times to all sides of these inequalities, we get

$$0 \leq \left[\max \overline{\mathrm{T}}^n h - \max_{y \in \mathcal{R}_\rightarrow} \overline{\mathrm{T}}^n h(y)\right] \underline{\mathrm{T}}^{n_o} I_{\mathcal{R}_\rightarrow}(x_o) \leq \max \overline{\mathrm{T}}^n h - \overline{\mathrm{T}}^{n_o+n} h(x_o) \tag{A.24}$$

after repeated use of properties $\overline{\mathrm{T}}_5$, $\overline{\mathrm{T}}_4$, $\overline{\mathrm{T}}_3$, and $\overline{\mathrm{T}}_7$, some rearranging, and evaluating in x_o. If we now take the limit inferior for $n \to \infty$ of all sides in these inequalities, we find

$$0 \leq \underline{\mathrm{T}}^{n_o} I_{\mathcal{R}_\rightarrow}(x_o) \liminf_{n\to\infty} \left[\max \overline{\mathrm{T}}^n h - \max_{y \in \mathcal{R}_\rightarrow} \overline{\mathrm{T}}^n h(y)\right] \leq \liminf_{n\to\infty} \left[\max \overline{\mathrm{T}}^n h - \overline{\mathrm{T}}^{n_o+n} h(x_o)\right]. \tag{A.25}$$

Since $\max \overline{\mathrm{T}}^n h \to L(h)$ and $\max_{y \in \mathcal{R}_\rightarrow} \overline{\mathrm{T}}^n h(y) \to l(h)$ [by the reasoning above, $\overline{\mathrm{T}}^n h(y) \to l(h)$ for all $y \in \mathcal{R}_\rightarrow$], we infer that $\liminf_{n\to\infty} \left[\max \overline{\mathrm{T}}^n h - \max_{y \in \mathcal{R}_\rightarrow} \overline{\mathrm{T}}^n h(y)\right] = L(h) - l(h)$ from the properties of the lim inf operator . It also follows for similar reasons that

$$\liminf_{n\to\infty} \left[\max \overline{\mathrm{T}}^n h - \overline{\mathrm{T}}^{n_o+n} h(x_o)\right] = \lim_{n\to\infty} \max \overline{\mathrm{T}}^n h - \limsup_{n\to\infty} \overline{\mathrm{T}}^{n_o+n} h(x_o) = L(h) - L(h). \tag{A.26}$$

So we infer from Eq. (A.25) that $\underline{\mathrm{T}}^{n_o} I_{\mathcal{R}_\rightarrow}(x_o)[L(h) - l(h)] = 0$ and, therefore, that also in this case, $l(h) = L(h)$, since, by assumption, $\underline{\mathrm{T}}^{n_o} I_{\mathcal{R}_\rightarrow}(x_o) > 0$. ■

Proof of Theorem 5.1: Since $I_\mathcal{X} \min \overline{\mathrm{T}}^n h \leq \overline{\mathrm{T}}^n h \leq I_\mathcal{X} \max \overline{\mathrm{T}}^n h$ and by Lemma A.5, both sequences $\min \overline{\mathrm{T}}^n h$ and $\max \overline{\mathrm{T}}^n h$ converge to the same real limit, which we denote by μ_h, it follows that $\overline{\mathrm{T}}^n h$ converges (pointwise) to $I_\mathcal{X} \mu_h$: $\lim_{n\to\infty} \overline{\mathrm{T}}^n h = I_\mathcal{X} \mu_h$. If we use the continuity of the upper expectation operator $\overline{E}_1$, as well as properties $\overline{\mathrm{T}}_4$ and $\overline{\mathrm{T}}_3$, we get

$$\lim_{n\to\infty} \overline{E}_1(\overline{\mathrm{T}}^{n-1} h) = \overline{E}_1\left(\lim_{n\to\infty} \overline{\mathrm{T}}^{n-1} h\right) = \overline{E}_1(I_\mathcal{X} \mu_h) = \mu_h, \tag{A.27}$$

and this limit is indeed independent of the choice of $\overline{E}_1$. Hence, we find for the limit that $\overline{E}_\infty(h) = \mu_h$.

To complete the proof, consider any upper expectation $\overline{E}_1$ on $\mathcal{L}(\mathcal{X})$ and any h in $\mathcal{L}(\mathcal{X})$; then for all $n \in \mathbb{N}$, $\overline{E}_1(\overline{\mathrm{T}}^n h) = \overline{E}_1(\overline{\mathrm{T}}^{n-1} \overline{\mathrm{T}} h)$. If we let $n \to \infty$ on both sides of this equality, we find that $\overline{E}_\infty(h) = \overline{E}_\infty(\overline{\mathrm{T}} h)$, showing that $\overline{E}_\infty$ is indeed $\overline{\mathrm{T}}$-invariant. Now, let $\overline{E}_\mathrm{i}$ be any $\overline{\mathrm{T}}$-invariant upper expectation on $\mathcal{L}(\mathcal{X})$. Then we find for any h in $\mathcal{L}(\mathcal{X})$ and for all $n \in \mathbb{N}$ that $\overline{E}_\mathrm{i}(\overline{\mathrm{T}}^{n-1} h) = \overline{E}_\mathrm{i}(h)$, and if we let $n \to \infty$ on both sides of this equality, we find that $\overline{E}_\infty(h) = \overline{E}_\mathrm{i}(h)$. ■

Proof of Proposition 5.2: We begin with the first statement. It clearly suffices to prove that for any $k \in \mathbb{N}$, with obvious notations, $\mathcal{T}_{\overline{\mathrm{T}}} \cdot \mathcal{T}_{\overline{\mathrm{T}}^k} \subseteq \mathcal{T}_{\overline{\mathrm{T}}^{k+1}}$. In other words, consider any $R \in \mathcal{T}_{\overline{\mathrm{T}}}$ and any $S \in \mathcal{T}_{\overline{\mathrm{T}}^k}$; then we have to show that $T := RS \in \mathcal{T}_{\overline{\mathrm{T}}^{k+1}}$. By Eq. (42), $R \in \mathcal{T}_{\overline{\mathrm{T}}}$ means that for all $x \in \mathcal{X}$, there is some $r(\cdot|x) \in \mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$ such that $R_{xy} = r(y|x)$ for all $y \in \mathcal{X}$. Similarly, by

Eq. (42), $S \in \mathcal{T}_{\overline{\mathrm{T}}^k}$ means that for all $y \in \mathcal{X}$, there is some $s(\cdot|y) \in \mathcal{Q}_{\overline{\mathrm{T}}^k}(\cdot|y)$ such that $S_{yz} = r(z|y)$ for all $z \in \mathcal{X}$. Now, for all $x \in \mathcal{X}$ and all $h \in \mathcal{L}(\mathcal{X})$,

$$\begin{aligned}\overline{\mathrm{T}}^{k+1}h(x) &= \overline{\mathrm{T}}(\overline{\mathrm{T}}^k h)(x) \\ &\geq E_{r(\cdot|x)}(\overline{\mathrm{T}}^k h) = \sum_{y\in\mathcal{X}} r(y|x)\overline{\mathrm{T}}^k h(y) \\ &\geq \sum_{y\in\mathcal{X}} r(y|x)E_{s(\cdot|y)}(h) = \sum_{y\in\mathcal{X}} r(y|x)\sum_{z\in\mathcal{X}} s(z|y)h(z) = \sum_{y,z\in\mathcal{X}} R_{xy}S_{yz}h(z) \\ &= \sum_{z\in\mathcal{X}} T_{xz}h(z),\end{aligned}$$

where both inequalities follow from Eq. (40). If we now consider, for each $x \in \mathcal{X}$, the mass function $q(\cdot|x)$ given by $q(z|x) := T_{xz} = \sum_{y\in\mathcal{X}} s(z|y)r(y|x)$ for all $z \in \mathcal{X}$, then this means that $\overline{\mathrm{T}}^{k+1}h(x) \geq E_{q(\cdot|x)}(h)$ for all $h \in \mathcal{L}(\mathcal{X})$ and, therefore, $q(\cdot|x) \in \mathcal{Q}_{\overline{\mathrm{T}}^{k+1}}(\cdot|x)$, for all $x \in \mathcal{X}$, by Eq. (40). Hence, indeed, $T \in \mathcal{T}_{\overline{\mathrm{T}}^{k+1}}$, by Eq. (42).

On to the second statement. We give a proof by induction. We first show that the statement holds for $n = 1$. We know from the definition (40) of $\mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$ and Eq. (41) that for each x in $\mathcal{X}$, there is some $q(\cdot|x) \in \mathcal{Q}_{\overline{\mathrm{T}}}(\cdot|x)$ such that $\overline{\mathrm{T}}h(x) = \sum_{y\in\mathcal{X}} q(y|x)h(y)$. Therefore, the transition matrix T, defined by $T_{xy} := q(y|x)$ for all $x, y \in \mathcal{X}$, belongs to $\mathcal{T}_{\overline{\mathrm{T}}}$ [see Eq. (42)] and satisfies $\overline{\mathrm{T}}h(x) = \sum_{y\in\mathcal{X}} T_{xy}h(y) = (Th)_x$.

Next, we show that if the statement holds for $n = m$ [the induction hypothesis], it also holds for $n = m + 1$, where $m \in \mathbb{N}$. Consider the real-valued map $g := \overline{\mathrm{T}}^m h$; then $\overline{\mathrm{T}}^{m+1}h = \overline{\mathrm{T}}g$. We know from the above reasoning that there is some $T_1 \in \mathcal{T}_{\overline{\mathrm{T}}}$ such that $\overline{\mathrm{T}}g(x) = (T_1 g)_x$ for all $x \in \mathcal{X}$. Additionally, the induction hypothesis tells us that there is some $T_2 \in \mathcal{T}_{\overline{\mathrm{T}}}^m$ such that $g(y) = \overline{\mathrm{T}}^m h(y) = (T_2 h)_y$ for all $y \in \mathcal{X}$. Hence, we find that for all $x \in \mathcal{X}$,

$$\begin{aligned}\overline{\mathrm{T}}^{m+1}h(x) = \overline{\mathrm{T}}g(x) &= \sum_{y\in\mathcal{X}} (T_1)_{xy}g(y) \\ &= \sum_{y\in\mathcal{X}} (T_1)_{xy}\sum_{z\in\mathcal{X}} (T_2)_{yz}h(z) = \sum_{z\in\mathcal{X}} (T_1T_2)_{xz}h(z) = (T_1T_2h)_x, \qquad \textbf{(A.28)}\end{aligned}$$

and, clearly, $T_1T_2 \in \mathcal{T}_{\overline{\mathrm{T}}}^{m+1}$. This concludes the proof of the second statement.

The third statement is an immediate consequence of the first and second statements. ■

Finally, we turn to the proof of Proposition 5.3. We first prove an alternative characterization of the product scrambling property.

LEMMA A.6: *A set $\mathcal{T}$ of transition matrixes is product scrambling if and only if*

$$(\exists n \in \mathbb{N})(\forall k \geq n)(\forall T \in \mathcal{T}^k)(\forall x, y \in \mathcal{X})(\exists z \in \mathcal{X})T_{xz} > 0 \wedge T_{yz} > 0. \qquad \textbf{(A.29)}$$

PROOF: Recall that $\mathcal{T}$ is called product scrambling if

$$(\exists n \in \mathbb{N})(\forall T \in \mathcal{T}^n)\tau(T) < 1. \qquad \textbf{(A.30)}$$

Since the coefficient of ergodicity satisfies the submultiplicative property [14, Sect. 1.2]:

$$\tau(T_1T_2) \leq \tau(T_1)\tau(T_2) \text{ for all transition matrixes } T_1 \text{ and } T_2, \quad \textbf{(A.31)}$$

we see that the product scrambling condition is equivalent to (see also [14, Lemma 3.2] for a related result):

$$(\exists n \in \mathbb{N})(\forall k \geq n)(\forall T \in \mathcal{T}^k)\tau(T) < 1. \quad \textbf{(A.32)}$$

Now, use Eq. (46). ∎

PROOF OF PROPOSITION 5.3: Assume that $\mathcal{T}_{\overline{\mathrm{T}}}$ is product scrambling. We prove that this implies that the corresponding stationary imprecise Markov chain with upper transition operator $\overline{\mathrm{T}}$ is regularly absorbing: (1) It is top class regular and (2) for every y not in the top class $\mathcal{R}_{\rightarrow}$, there is some $n \in \mathbb{N}$ such that $\underline{\mathrm{T}}^n I_{\mathcal{R}_{\rightarrow}}(y) > 0$.

We first prove that the Markov chain has a top class under $\rightarrow$. It follows from the characterization (A.29) of the product scrambling condition in Lemma A.6 that

$$(\forall x, y \in \mathcal{X})(\exists z \in \mathcal{X})x \rightarrow z \wedge y \rightarrow z \quad \textbf{(A.33)}$$

if we also take into account Proposition 5.2. For any $x, y \in C$, where $C \subseteq \mathcal{X}$ is the [always nonempty] set of all maximal states, we know that $x \rightarrow z \Rightarrow z \rightarrow x$ and $y \rightarrow z \Rightarrow z \rightarrow y$ for all $z \in \mathcal{X}$, so we infer from Eq. (A.33) that both $x \rightarrow y$ and $y \rightarrow x$, so x and y communicate. This means that the whole of C forms one single communication class: C is the top class.

We now show that this top class C is regular (i.e., consists of a single cyclic subclass) if we recall our discussion of periodicity in Section 4.1. Let d_C be the period of the top class C and consider any x and y in C. Using the same reasoning as earlier, we infer from Eq. (A.29) and Proposition 5.2 that for large enough k,

$$(\exists z_k \in C)x \stackrel{k}{\rightarrow} z_k \wedge y \stackrel{k}{\rightarrow} z_k \quad \textbf{(A.34)}$$

(that $z_k \in C$ follows from the fact that x and y are maximal). Moreover, Proposition 4.1 tells us that for large enough ℓ and ℓ', $t_{z_kx} + \ell d_C \in N_{z_kx}$ and $t_{z_ky} + \ell' d_C \in N_{z_ky}$ and, therefore, also $k + t_{z_kx} + \ell d_C \in N_{xx}$ and $k + t_{z_ky} + \ell' d_C \in N_{yy}$. This implies that $t_{z_kx} = t_{z_ky}$ and, therefore, $t_{xy} = 0$: x and y belong to the same cyclic class. This holds for all $x, y \in C$, so C consists of only one cyclic class (under $\rightarrow$). The top class C is, in other words, aperiodic and therefore regular. This proves (1).

To prove (2), assume the stationary imprecise Markov chain is top class regular but not regularly absorbing. We show that the set of transition matrixes $\mathcal{T}_{\overline{\mathrm{T}}}$ cannot be product scrambling. By Definition 4.1, we know that there is some $y_0 \in \mathcal{X} \setminus \mathcal{R}_{\rightarrow}$ such that $\underline{\mathrm{T}}^n I_{\mathcal{R}_{\rightarrow}}(y_0) = 0$ for all $n \in \mathbb{N}$. If we now also invoke Eq. (43) in Proposition 5.2, we see that for all $n \in \mathbb{N}$, there is some $T_n^* \in \mathcal{T}_{\overline{\mathrm{T}}}^n$ such that

$$(\forall u \in \mathcal{R}_{\rightarrow})(T_n^*)_{y_0u} = 0. \quad \textbf{(A.35)}$$

Now, consider any x_0 in the top class $\mathcal{R}_{\rightarrow}$ [this is possible since by assumption $\mathcal{R}_{\rightarrow} \neq \emptyset$]. Since x_0 cannot communicate with any element outside $\mathcal{R}_{\rightarrow}$, we infer in particular from Eq. (43) in Proposition 5.2 that for all $n \in \mathbb{N}$,

$$(\forall v \in \mathcal{X} \setminus \mathcal{R}_{\rightarrow})(T_n^*)_{x_0v} = 0. \quad \textbf{(A.36)}$$

However, Eqs (A.35) and (A.36) taken together imply [see Eq. (46)] that $\tau(T_n^*) = 1$ for all $n \in \mathbb{N}$, so the set $\mathcal{T}_{\overline{\mathrm{T}}}$ is not product scrambling. ∎