Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 381 (2004) 61–76
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

A limit theorem for sets of stochastic matrices

Anne Condon [a,1], Michael Saks [b,*,2]

[a] Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, BC, Canada V6T1Z4

[b] Department of Mathematics–Hill Center, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA

Received 19 August 2003; accepted 18 September 2003

Submitted by R.A. Brualdi

Abstract

The following fact about (row) stochastic matrices is an easy consequence of well known results: for each positive integer $n \geqslant 1$ there is a positive integer $q = q(n)$ with the property that if A is any $n \times n$ stochastic matrix then the sequence of matrices $A^q, A^{2q}, A^{3q}, \ldots$ converges. We prove a generalization of this for sets of stochastic matrices under the Hausdorff metric. Let d be any metric inducing the standard topology on the set of $n \times n$ real matrices. For a matrix A and set of matrices $\mathscr{B}$ define $d(A, \mathscr{B})$ to be the infimum of $d(A, B)$ over all $B \in \mathscr{B}$. For two sets of matrices $\mathscr{A}$ and $\mathscr{B}$, define $d^+(\mathscr{A}, \mathscr{B})$ to be the supremum of $d(A, \mathscr{B})$ over all $A \in \mathscr{A}$, and define $d(\mathscr{A}, \mathscr{B})$ to be the maximum of $d^+(\mathscr{A}, \mathscr{B})$ and $d^+(\mathscr{B}, \mathscr{A})$. This is the Hausdorff metric on the set of subsets of $n \times n$ stochastic matrices. If $\mathscr{A}$ is a set of stochastic matrices and k is a positive integer, define $\mathscr{A}^{(k)}$ to be the set of all matrices expressible as a product of a sequence of k matrices from $\mathscr{A}$. We prove: For each positive integer n there is a positive integer $p = p(n)$ such that if $\mathscr{A}$ is any subset of $n \times n$ stochastic matrices then the sequence of subsets $\mathscr{A}^{(p)}, \mathscr{A}^{(2p)}, \mathscr{A}^{(3p)}, \ldots$ converges with respect to the Hausdorff metric.

PACS: 15A51

Keywords: Stochastic matrices; Markov set chains; Hausdorff metric

* Corresponding author.
E-mail addresses: condon@cs.ubc.ca (A. Condon), saks@math.rutgers.edu (M. Saks).

[1] Supported in part by an NSERC grant.
[2] Supported in part by NSF under grants CCR-9700239 and CCR-9988526.

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2003.09.020

1. Introduction

A fundamental fact about (finite-state, discrete-time) Markov chains is that if all transition probabilities are nonzero, then the chain has a limiting distribution. This reflects the fact that the state transition matrix A associated to any Markov chain belongs to the set $\mathscr{S}_n$ of (row)-stochastic matrices (nonnegative matrices with each row sum equal to 1) and thus has largest eigenvalue 1. If all entries are nonzero the eigenspace for 1 has a unique (row) eigenvector v with entries summing to 1, and thus the sequence of powers $A^1, A^2, \ldots$ converges to the matrix whose rows are each equal to v.

For arbitrary stochastic matrices, the powers $A^1, A^2, \ldots$ need not converge. However, it can be shown that there is a number $p = p(n)$ such that for any $A \in \mathscr{S}_n$, A^p is a block diagonal matrix each of whose blocks has no zero entries. Thus by the previous fact, the sequence $\{A^{jp} : j \geqslant 0\}$ converges, i.e., the sequence $\{A^j : j \geqslant 1\}$ approaches a *periodic limit* with period dividing $p(n)$.

We consider the more general situation of a discrete-time process on n states whose possible behaviors are characterized by an arbitrary *subset* $\mathscr{A}$ of $\mathscr{S}_n$. At each step the process makes a state transition according to one of the matrices A in $\mathscr{A}$. The evolution of the system is thus described by a sequence $A_1, A_2, \ldots$ of matrices each from $\mathscr{A}$, and each such sequence corresponds to a possible behavior of the system. For $k \geqslant 1$, let $\mathscr{A}^k$ denote the set of all sequences $(A_1, \ldots, A_k)$ from $\mathscr{A}$ and let $\mathscr{A}^{(k)}$ denote the set of all products of the form $A_1 \ldots A_k$ where $A_i \in \mathscr{A}$. We view the subsets of $\mathscr{S}_n$ as points of a metric space under the Hausdorff metric (see Section 2.1). Our main result is:

Theorem 1.1. *For each natural number n there is a natural number $p = p(n)$, such that if $\mathscr{A} \subseteq \mathscr{S}_n$ then the sequence of sets $\{\mathscr{A}^{(pi)} : i \geqslant 1\}$ is convergent.*

Sequences of the form $\mathscr{A}^{(1)}, \mathscr{A}^{(2)}, \ldots$ where $\mathscr{A}$ is a compact set of stochastic matrices are called Markov set-chains. The explicit study of such sequences was initiated by Hartfiel (see [4,5] and the references therein), though as explained in [4] there are many related antecedents. Much of the previous research has been concerned with identifying sufficient conditions on the set $\mathscr{A}$ that ensure that the sequence $\{\mathscr{A}^{(i)} : i \geqslant 1\}$ is convergent.

Theorem 1.1 arose in connection with a problem in theory of computation posed by Dwork and Stockmeyer [2] concerning interactive finite automata: is it true that every language that admits an interactive proof of membership with a finite state verifier must be regular? (We only mention this problem in passing, and will not give definitions here.) Theorem 1.1 implies an affirmative answer in the very special case of unary languages where the verifier is restricted to one-way access to the input. This special case was already known [1] (by an easier argument), however, we think that Theorem 1.1 is of independent interest and also that there is a possibility of extending the ideas of Theorem 1.1 to handle the as yet unsolved case of unary languages where the verifier has two-way access to the input. We hope to explore this in a future paper.

2. Preliminaries

This section reviews various notions from point-set topology, combinatorics and the theory of stochastic matrices, and establishes various preliminary results.

2.1. Convergence with respect to the Hausdorff metric

We review basic definitions and results concerning the Hausdorff metric. In particular, Theorem 2.3(3) below will provide a useful sufficient condition for proving convergence. The facts cited here are elementary and standard but we do not know a reference that summarizes them in the form we need, so we provide proofs for them.

Let M be a metric space with distance function d. For $x \in M$, $\varepsilon > 0$, we write $B_\varepsilon(x)$ for the closed ball of radius ε around x. For $X \subseteq M$, $B_\varepsilon(X) = \cup_{x\in X} B_\varepsilon(x)$. For $X \subseteq M$ we write $\overline{X}$ for the closure of X.

For subsets X and Y, define $d^+(X, Y)$ to be the infimum of the set $\{\varepsilon | X \subseteq B_\varepsilon(Y)\}$ (the limit is ∞ if the set is empty). For a point x, we write $d^+(x, Y)$ for $d^+(\{x\}, Y)$. Note that $d^+(X, Y) = \sup_{x\in X} d^+(x, Y)$. It is easy to check: (i) $d^+(X, Z) \leqslant d^+(X, Y) + d^+(Y, Z)$ and (ii) $d^+(X, Y) = 0$ if and only if $\overline{X} \subseteq \overline{Y}$.

Define $d(X, Y)$ to be the maximum of $d^+(X, Y)$ and $d^+(Y, X)$. Then d satisfies (i) $d(X, Z) \leqslant d(X, Y) + d(Y, Z)$, (ii) $d(X, Y) = d(Y, X)$ and (iii) $d(X, Y) = 0$ if and only if $\overline{X} = \overline{Y}$. d is generally not a metric on the power set of M but it is a metric when restricted to the set of compact subsets of M.

A set X is an *upper limit* for $\{X_i\}$ if $\{d^+(X_i, X) : i \geqslant 1\}$ converges to 0 and is a *lower limit* if $\{d^+(X, X_i) : i \geqslant 1\}$ converges to 0, and is a *limit* if it is both an upper limit and a lower limit.

A point x is a *strong limit point* for $\{X_i\}$ if every neighborhood of x intersects all but finitely many of the X_i and is a *weak limit point* if every neighborhood of x intersects infinitely many of the X_i. Write X_{strong} and X_{weak} for the set of weak and strong limit points. It is easy to see that both of these sets are closed. Trivially, $X_{\text{strong}} \subseteq X_{\text{weak}}$ and we will say that the sequence $\{X_i\}$ is *regular* if $X_{\text{weak}} = X_{\text{strong}}$.

Proposition 2.1. *Let $\{X_i\}$ be a sequence of subsets of M.*

1. *If Y is a lower limit for $\{X_i\}$ then $Y \subseteq X_{\text{strong}}$.*
2. *If Y is an upper limit for $\{X_i\}$ then $X_{\text{weak}} \subseteq \overline{Y}$.*

Proof. If Y is a lower limit for $\{X_i\}$, then $d^+(Y, X_i)$ tends to 0, and so for any $y \in Y$, $d^+(y, X_i)$ tends to 0, which implies that any neighborhood of y intersects all but finitely many of the X_i; i.e., y is a strong limit point.

Next suppose that Y is an upper limit for $\{X_i\}$ and let y be a weak limit point. We want to show that y is in the closure of Y, so we fix $\varepsilon > 0$ and show that $d(y, Y) \leqslant \varepsilon$. There is an i_0 such that $d^+(X_i, Y) \leqslant \varepsilon/2$ for $i \geqslant i_0$. Also the $\varepsilon/2$ neighborhood of y intersects infinitely many X_i so it contains some point $z \in X_j$ with $j \geqslant i_0$. Then $d(y, Y) \leqslant d(y, z) + d(z, Y) \leqslant d(y, z) + d(X_j, Y) \leqslant \varepsilon$. □

We will say that $\{X_i\}$ is *lower convergent* if X_{strong} is a lower limit, *upper convergent* if X_{weak} is an upper limit, and *convergent* if it has a limit.

Proposition 2.2. *If $\{X_i\}$ is convergent then it is regular and the set $X_{\text{strong}} = X_{\text{weak}}$ is the unique closed set that is a limit for $\{X_i\}$.*

Proof. Suppose $\{X_i\}$ is convergent with limit Y, and assume without loss of generality that Y is closed. By Proposition 2.1, $X_{\text{weak}} \subseteq Y \subseteq X_{\text{strong}}$, since $X_{\text{strong}} \subseteq X_{\text{weak}}$ we conclude $X_{\text{weak}} = Y = X_{\text{strong}}$. □

A sequence $\{X_i\}$ of subsets is *forward Cauchy* if for all $\varepsilon > 0$ there exists an $m_0 = m_0(\varepsilon)$ such that for m_1, m_2 satisfying $m_2 \geqslant m_1 \geqslant m_0$, $d^+(X_{m_1}, X_{m_2}) \leqslant \varepsilon$. It is *backward Cauchy* if for all $\varepsilon > 0$ there exists an m_0 such that for m_1, m_2 satisfying $m_2 \geqslant m_1 \geqslant m_0$, $d^+(X_{m_2}, X_{m_1}) \leqslant \varepsilon$ and it is *Cauchy* if it is both forward and backward Cauchy. It is easy to see that a convergent sequence $\{X_i\}$ is Cauchy.

If we restrict to compact metric spaces, we get some nice implications.

Theorem 2.3. *Let $\{X_i\}$ be a sequence of sets in a compact metric space. Then*

1. $\{X_i\}$ *is upper convergent.*
2. $\{X_i\}$ *is lower convergent.*
3. *If $\{X_i\}$ is either forward or backward Cauchy then it is convergent.*

Proof. Suppose that $\{X_i\}$ is not upper convergent. Then there is an $\varepsilon > 0$, a sequence of indices $i_1 < i_2 < \cdots$, a sequence $\{x_{i_j} \in X_{i_j}\}$ such that $x_{i_j} \notin B_\varepsilon(X_{\text{weak}})$. By compactness, the sequence $\{x_{i_j}\}$ has an accumulation point z which by definition belongs to X_{weak}. Then $B_\varepsilon(z)$ contains at least one (in fact, infinitely many) of the x_{i_j} contradicting that for all j, $x_{i_j} \notin B_\varepsilon(X_{\text{weak}})$.

Suppose that $\{X_i\}$ is not lower convergent. Then there is an $\varepsilon > 0$, an infinite sequence of indices $i_1 < i_2 < \cdots$, and a sequence $\{x_{i_j} \in X_{\text{strong}}\}$ such that $x_{i_j} \notin B_\varepsilon(X_{i_j})$. By compactness, the sequence $\{x_{i_j}\}$ has an accumulation point z, which is in X_{strong} since X_{strong} is closed. Then $z \in B_{\varepsilon/2}(X_{i_j})$ for all but finitely many j (by the definition of X_{strong}) and $d(x_{i_j}, z) \leqslant \varepsilon/2$ for all but finitely many j so $x_{i_j} \in B_\varepsilon(X_{i_j})$ for all but finitely many j, a contradiction.

Now suppose that $\{X_i\}$ is forward Cauchy or backward Cauchy. To show that $\{X_i\}$ is convergent, suppose $z \in X_{\text{weak}}$, we show that $z \in X_{\text{strong}}$. Fix $\varepsilon > 0$, we show that $B_\varepsilon(z)$ intersects all but finitely many of the X_i, or equivalently $z \in B_\varepsilon(X_i)$ for all but finitely many i. Since $z \in X_{\text{weak}}$ there is an infinite sequence $i_1 < i_2 < \cdots$ such that $z \in B_{\varepsilon/2}(X_{i_j})$ for all j. Since $\{X_i\}$ is forward Cauchy or backward Cauchy, we can choose an index $m_0 = m_0(\varepsilon/2)$ as in the definition and we may assume that $m_0 \geqslant i_1$. We claim that $z \in B_\varepsilon(X_i)$ for all $i > m_0$ and hence $z \in X_{\text{strong}}$. To see the claim, let $i > m_0$. If $i = i_j$ for some j the claim is trivial; otherwise we choose a j so that $i_j < i < i_{j+1}$. If $\{X_i\}$ is forward Cauchy then $B_\varepsilon(X_i)$ contains $B_{\varepsilon/2}(X_{i_j})$ and if it is backward Cauchy $B_\varepsilon(X_i)$ contains $B_{\varepsilon/2}(X_{i_{j+1}})$. In either case, $z \in B_\varepsilon(X_i)$, as claimed. □

2.2. *Intervals and interval chains*

If x, y are real numbers between 0 and 1, interval notation has the usual meaning. If x, y are nonnegative integers then interval notation is used to denote subsets of integers, e.g. $[x, y]$ is the set of integers $\{x, x+1, \dots, y-1, y\}$. (This means that $[0, 1]$ is potentially ambiguous, but the meaning will be clear from the context.) We write $[n]$ for $[1, n]$. For integers a, b, we write $Int[a, b]$ for the set of (integer) intervals contained in $[a, b]$, and $Int[m] = Int[1, m]$. An *interval chain* σ is a sequence $(\sigma_1, \sigma_2, \dots, \sigma_k)$ where each σ_i is an interval $[l_i, r_i]$ and $l_i = r_{i-1} + 1$ for each $i \in [2, k]$. The number k is the length of the interval chain. We say that the chain *ends at* r_k and *spans* the interval $[l_1, r_k]$. For example $([2, 4], [5, 5], [6, 9])$ is an interval chain of length 3 that ends at 9 and spans $[2, 9]$.

We will have need to consider functions mapping $Int[m]$ to a finite set C; we call such a map a C-coloring of $Int[m]$. The following lemma is essentially due to Erdös and Szekeres [3]:

Lemma 2.4. *Let C be a finite set and h a positive integer. If $m \geqslant h^{|C|}$ then given any C-coloring of $Int[m]$ there is an interval chain of length h that is monochromatic, i.e. in which all parts get the same color.*

Proof. For each integer $j \in [m]$ we define a function f_j on the set C of colors where, for $c \in C$, $f_j(c)$ is the maximum number of parts of an interval chain ending at j all of whose parts are color c. We claim that the functions f_i and f_j are different for all $i \neq j$. Assume $i > j$ and let c be the color of $[j+1, i]$. Then the interval chain ending at $[j]$ of length $f_j(c)$ having all parts of color c can be augmented by $[j+1, i]$ to get an interval chain ending at i of length $f_j(c) + 1$ having all parts of color c. Hence $f_i(c) > f_j(c)$ and we conclude that the functions f_i for $i \in [0, m]$ are distinct. Since there are $(h-1)^{|C|} < m$ functions from C to $[h-1]$, the pigeonhole principle implies that there is an index $m' \leqslant m$ and a color c such that $f_{m'}(c) \geqslant h$. Thus there is a monochromatic interval chain with h parts. □

2.3. Partial partitions

If n is an integer, a *partial partition* is a family Π of pairwise disjoint subsets of $[n]$. For a partial partition Π we write $\cup\Pi$ for $\cup_{S\in\Pi}S$, and $Res(\Pi)$, the *residue* of Π, is $[n]-\cup\Pi$. For $j\in\cup\Pi$, $\Pi[s]$ denotes the unique set in Π that contains s.

2.4. Directed graphs

For our purposes a directed graph D on vertex set $[n]$ is a subset of $[n]\times[n]$. An element (a,b) of the graph is an *arc* with *source* a and *target* b. An arc of the form (s,s) is a *loop*. Since our digraphs arise as state spaces for finite Markov chains, we refer to vertices as *states*. If $(s,t)\in D$, we say that t is *accessible* from s. A state s is *self-accessible* if (s,s) is an arc. $D^+(s)$ is the set of states accessible from s.

A *walk* of length $k\geqslant 1$ in D from state s to state t is a sequence $s=s_0,s_1,s_2,\ldots,s_{k-1},s_k=t$ of (not necessarily distinct) states such that $(s_0,s_1),(s_1,s_2),\ldots,(s_{k-1},s_k)\in D$. We say that t is *reachable* from s provided that there is a walk from s to t. We say s is *self-reachable* if there is a walk of length at least 1 from s to itself.

A subset S of states is *absorbing* with respect to D if there are no arcs from S to $[n]-S$. The intersection of absorbing sets is absorbing and hence any two minimal absorbing sets (under containment) are disjoint. The collection Γ_D of minimal absorbing sets is a partial partition of $[n]$. States belonging to $\cup\Gamma_D$ are said to be *recurrent* and recur_D is the set of recurrent states. States not in $\cup\Gamma_D$ are said to be *transient* and trans_D is the set of transient states.

To each partial partition Π of $[n]$, we associate a digraph $G(\Pi)=(\cup_{S\in\Pi}S\times S)\cup(Res(\Pi)\times[n])$. $G(\Pi)$ is the unique maximum digraph (under containment) for which $\Gamma_{G(\Pi)}=\Pi$.

The Boolean product of two digraphs D_1 and D_2 is the digraph $D_1D_2=\{(s,t):\exists u,(s,u)\in D_1,(u,t)\in D_2\}$. The Boolean power of a digraph D^k is defined in the obvious way. It is easy to see that $(s,t)\in D^k$ if and only if there is a walk of length k from s to t in D.

A digraph D is said to be

- *admissible* if each vertex is the source of some arc (possibly an arc to itself);
- S*-avoiding* for $S\subseteq[n]$ if no state in S is the target of any arc;
- Π*-structured* for a partial partition Π if $\Gamma_D=\Pi$;
- Π*-absorbing* for a partial partition Π if each $S\in\Pi$ is absorbing, equivalently, $D\subseteq G(\Pi)$;
- *stable* if $D=D^2$.

Lemma 2.5. *For any digraph D on vertex set $[n]$, $D^{n!}$ is stable.*

Proof. First we show:

Claim. Let F be a digraph on $[n]$ having at least one self-reachable state and having the property that every self-reachable state is self-accessible. Let $i \geqslant n-1$. Then F^i is stable.

To prove the claim, let F be a digraph satisfying the hypothesis and $i \geqslant n-1$. Let $(s,t) \in [n] \times [n]$. We need that $(s,t) \in F^i$ if and only if there is a $u \in [n]$ with $(s,u), (u,t) \in F^i$.

If $(s,t) \in F^i$ there is a walk from s to t of length i in F. Since $i \geqslant n-1$, either this walk repeats some state or it contains all n states. In either case, the walk contains at least one state u that is self-accessible in F (since F has at least one self-reachable state and every self-reachable state is self-accessible). Then $(s,u), (u,t) \in F^i$. Conversely, suppose there is a $u \in [n]$ such that $(s,u), (u,t) \in F^i$. Then there is a walk W of length $2i$ from s to t, and as above, it contains a self-accessible state. Among all walks from s to t that contain a self-accessible state, choose a shortest one. This walk has no repeated state, since if s' is repeated, we may shorten the walk, and the shortened walk still contains s', which is self-reachable and hence self-accessible. Hence the length j of the walk is at most $n-1$, and we may then lengthen the walk to exactly i by inserting $i-j$ occurences of some self-accessible state after its occurence in the original walk. Thus $(s,t) \in F^i$.

Using the claim, we prove the lemma. If D is acyclic (i.e., no vertex is self-reachable) then D^i is the empty graph for $i \geqslant n$, which is stable, so assume that D has at least one cycle. If D contains a cycle through all of the states, then all states in D^n are self-accessible, and setting $F = D^n$ and $i = (n-1)!$ in the claim, we conclude that $D^{n!}$ is stable. Otherwise, for each self-reachable state s of D, the length l_s of the shortest cycle containing s is less than n and hence divides $(n-1)!$ Thus every self-reachable state of D (and hence also of $D^{(n-1)!}$) is self-accessible in $D^{(n-1)!}$. Now apply the claim with $F = D^{(n-1)!}$ and $i = n$. $\square$

2.5. Matrices

For a matrix A, $\mu(A)$ denotes the least absolute value of any nonzero entry of A. For two matrices A and B, $A \leqslant B$ means $A(s,t) \leqslant B(s,t)$ for all $s, t \in [n] \times [n]$.

The norm of a matrix, $\|A\|$ is the maximum over $s \in [n]$ of $\sum_{t \in [n]} |A(s,t)|$. The distance between two matrices $d(A,B) = \|A-B\|$. We recall an elementary property of this norm:

Proposition 2.6. *Let $A_1, A_2, \ldots, A_m, B_1, B_2, \ldots, B_m$ be matrices each of norm at most* 1. *Then*:

$$d(A_1A_2\cdots A_m, B_1B_2\cdots B_m) \leqslant \sum_{i=1}^{m} d(A_i, B_i).$$

If $(A_1, A_2, \ldots, A_m)$ is a sequence of matrices, and $\sigma = [i, j]$ is an interval contained in $[m]$ then A_σ denotes the product $A_iA_{i+1}A_{i+2}\cdots A_j$. We call the sequence $(A_i, A_{i+1}, \ldots, A_j)$ a *segment* of $(A_1, \ldots, A_m)$.

2.6. Stochastic matrices, and Markov chains

We identify a stochastic matrix A with the Markov chain it defines. If $s \in [n]$, $T \subseteq [n]$ then $A(s, T)$ is defined to be $\sum_{t \in T} A(s, t)$ and is equal to the conditional probability that the chain enters a state in T at some step $i + 1$ given that it is at s at step i. A sequence $(A_1, A_2, \ldots, A_m)$ of stochastic matrices, corresponds to an m-step Markov process, where the transition probabilities of the ith step are given by A_i, and $A_{[1,m]}$ is the transition matrix for the entire process.

To each $n \times n$ stochastic matrix A is associated the directed graph $D(A)$ on $[n]$ with arc set $\{(s, t) : A(s, t) > 0\}$. Note that $D(AB) = D(A)D(B)$ where the product on the right is Boolean multiplication. We call $D(A)$ the *pattern* of A. Note that the pattern of a stochastic matrix is necessarily an admissible digraph, as defined earlier.

We will be concerned with a number of properties of A that only depend on $D(A)$. We adapt the terminology for digraphs to matrices and Markov chains. Thus, the terms accessible, reachable, Π-absorbing, S-avoiding, stable, etc. are defined for stochastic matrices A by referring to $D(A)$. In particular, the partial partition Γ_A is defined to be $\Gamma_{D(A)}$. Note that the terms recurrent and transient as defined for digraphs have the usual meaning for Markov chains: recurrent states are those that are visited infinitely often with positive probability, and transient states are states that are visited finitely often with probability 1.

As usual, a matrix A is *idempotent* if $A^2 = A$. We say that A is *quasi-idempotent* if the submatrix corresponding to the recurrent states is idempotent. The reader can prove:

Proposition 2.7. *A stochastic matrix A is quasi-idempotent if and only if for each $S \in \Gamma_A$ the rows of A corresponding to S are identical.*

Lemma 2.8 is critical to the main argument. It identifies some special conditions which guarantee that the product ABC of three stochastic matrices is independent of the middle matrix B.

Lemma 2.8. *Let Π be a partial partition of $[n]$ and $A, C \in \mathscr{S}_n$. If C is Π-structured and quasi-idempotent and A is $Res(\Pi)$-avoiding, then for any Π-absorbing $B \in \mathscr{S}_n$ the matrix ABC satisfies*:

$$ABC(s,t) = \begin{cases} 0 & t \in Res(\Pi), \\ A(s, \Pi[t])C(t,t) & t \in \cup\Pi. \end{cases}$$

In particular, the matrix ABC is independent of B.

Proof. We determine $ABC(s, t)$ by analyzing the three step stochastic process associated to (A, B, C). Suppose first that $t \in Res(\Pi)$. Starting from any state s, the first step of the process moves to a state in $\cup\Pi$ with probability 1, since A is $Res(\Pi)$-avoiding, and the next two steps keep the process in $\cup\Pi$. Hence $ABC(s, t) = 0$ as required.

Next suppose $t \in \cup\Pi$. Because C is Π-absorbing, the process can end in t only if it is in $\Pi[t]$ after the second step. Hence $ABC(s,t) = \sum_{u\in\Pi[t]} AB(s,u)C(u,t) = AB(s,\Pi[t])C(t,t)$ where the last equality comes from Proposition 2.7 and the fact that C is quasi-idempotent. Since A is $Res(\Pi)$-avoiding, the process is in $\cup\Pi$ after the first step, and since B is Π-absorbing, it stays in the same set of Π after the second step and hence $AB(s,\Pi[t]) = A(s,\Pi[t])$. Thus $ABC(s,t) = A(s,\Pi[t])C(t,t)$ as required. $\square$

2.7. Two operators on stochastic matrices

We will need two operators mapping $\mathscr{S}_n$ to $\mathscr{S}_n$. The first is defined in terms of a given digraph D, and maps A to a stochastic matrix $\widetilde{A}\langle D\rangle$ whose pattern is contained in D, and such that $\widetilde{A}\langle D\rangle$ is close to A:

$$\widetilde{A}\langle D\rangle(s,t) = \begin{cases} A(s,t) + \frac{A(s,[n]-D^+(s))}{|D^+(s)|} & (s,t) \in D, \\ 0 & (s,t) \notin D. \end{cases}$$

From the definition that $\widetilde{A}\langle D\rangle$ has pattern contained in D and has the same row sums as A (provided that D is admissible). To bound $d(A,\widetilde{A}\langle D\rangle)$, we note that for each s,

$$\begin{aligned}\sum_t |\widetilde{A}\langle D\rangle(s,t) - A(s,t)| &= \sum_{t\in D^+(s)} |\widetilde{A}\langle D\rangle(s,t) - A(s,t)| \\ &\quad + \sum_{t\notin D^+(s)} |\widetilde{A}\langle D\rangle(s,t) - A(s,t)| \\ &= 2A(s,[n]-D^+(s)).\end{aligned}$$

Summarizing the above, we have:

Proposition 2.9. *If A is a stochastic matrix and D is an admissible digraph then $\widetilde{A}\langle D\rangle$ is a stochastic matrix with pattern D and $d(A,\widetilde{A}\langle D\rangle) \leqslant 2\max_{s\in[n]} A(s,[n]-D^+(s))$.*

The second operator maps A to a quasi-idempotent matrix $\widehat{A}$ close to A, such that $\Gamma_{\widehat{A}} = \Gamma_A$. We define $\widehat{A}$ as follows: If $s \in \mathrm{trans}_A$ then row s of $\widehat{A}$ is equal to row s of A. If $s \in \mathrm{recur}_A$ then row s of $\widehat{A}$ is equal to the arithmetic average of the rows of A corresponding to $t \in \Gamma_A[s]$. Clearly $\widehat{A}$ is stochastic and $\Gamma_{\widehat{A}} = \Gamma_A$. All of the rows of $\widehat{A}$ corresponding to states in the same set of Γ_A are identical so, by Proposition 2.7, $\widehat{A}$ is quasi-idempotent.

Next we bound $d(\widehat{A},A)$. For $t \in \mathrm{recur}_A$, define $\max_A(t)$ (resp. $\min_A(t)$) to be the maximum (resp. minimum) of $A(s,t)$ over $s \in \Gamma_A[t]$, and define $\Delta_A(t) = \max_A(t) - \min_A(t)$. Let $\Delta_A = \max(\Delta_A(t) : t \in \mathrm{recur}_A)$. To upper bound $d(\widehat{A},A)$ it suffices to upper bound $\sum_t |A(s,t) - \widehat{A}(s,t)|$ for arbitrary $s \in [n]$. If $s \in \mathrm{trans}_A$ then this is 0.

Otherwise the only nonzero terms in the sum are those corresponding to $t \in \Gamma_A[s]$ and for those $\min_A(t) \leqslant A(s,t), \widehat{A}(s,t) \leqslant \max_A(t)$, from which we conclude that $|A(s,t) - \widehat{A}(s,t)| \leqslant \Delta_A(t)$. This implies that $\sum_t |A(s,t) - \widehat{A}(s,t)| \leqslant \Delta_A n$ and we have:

Proposition 2.10. *Let A be a stochastic matrix. Then $\widehat{A}$ is stochastic, Π_A-structured, and quasi-idempotent and $d(A, \widehat{A}) \leqslant \Delta_A n$.*

3. Proof of Theorem 1.1

Fix $\mathscr{A} \subseteq \mathscr{S}_n$. It suffices to prove the theorem in the case that $\mathscr{A}$ is closed. Also, since $\mathscr{S}_n$ is compact, it suffices, by Theorem 2.3, to show that $\{\mathscr{A}^{(pi)} : i \geqslant 1\}$ is forward Cauchy. This is equivalent to showing:

Lemma 3.1. *For each natural number n there is a natural number $p = p(n)$ with the following property: Let $\mathscr{A}$ be a closed subset of $\mathscr{S}_n$. For each $\varepsilon > 0$, there is an integer $m_0 = m_0(\mathscr{A}, \varepsilon)$ such that if $m \geqslant m_0$ and $A \in \mathscr{A}^{(m)}$, then for any positive integer i there is a matrix $C_i \in \mathscr{A}^{(m+ip)}$ such that $\|C_i - A\| \leqslant \varepsilon$.*

Given ε, we will choose $m_0 = m_0(\mathscr{A}, \varepsilon)$ sufficiently large. We are then given an arbitrary sequence $(A_1, A_2, \ldots, A_m)$ from $\mathscr{A}^m$ with $m \geqslant m_0$ and must show that for some p depending only on n, and for any $i \geqslant 1$ there is a sequence $(B_1, B_2, \ldots, B_{m+ip})$ of matrices from $\mathscr{A}^{m+ip}$ such that $\|B_1 B_2 \cdots B_{m+ip} - A_1 A_2 \Gamma A_m\| \leqslant \varepsilon$.

We will use Lemma 2.8. Lemma 3.2 below, asserts that we can partition any long enough sequence of matrices into five segments so that, denoting by P_i the product of the ith segment, we have that for some partial partition Π, P_2 is very close to a $Res(\Pi)$-avoiding matrix, P_3 is very close to a Π-absorbing matrix and P_4 is very close to a Π-structured quasi-idempotent matrix. Now Lemma 2.8 will imply that if we replace P_3 by any matrix N whose product is close to some Π-absorbing matrix, then $P_1 P_2 N P_4 P_5$ is close to $P_1 P_2 P_3 P_4 P_5$. So it suffices to show that if k is the length of the third segment, then for some p depending on n and for each $i \geqslant 1$, we can find a matrix $N_i \in \mathscr{A}^{(k+ip)}$ that is close to Π-absorbing. This (or something like it) will follow from Lemma 3.3.

We now formulate the two main lemmas, and show that they imply Lemma 3.1.

Lemma 3.2. *Let n be a positive integer and $\varepsilon' > 0$. There is an integer $b = b(n, \varepsilon')$ such that if $(B_1, B_2, \ldots, B_b)$ is any sequence of $n \times n$ stochastic matrices then there exists a partial partition Π of $[n]$ and an interval chain $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ that spans $[b]$ satisfying:*

1. *There is a $Res(\Pi)$-avoiding matrix L_2 with $d(L_2, B_{\sigma_2}) \leqslant \varepsilon'$.*
2. *There is a Π-absorbing matrix L_3 with $d(L_3, B_{\sigma_3}) \leqslant \varepsilon'$.*
3. *There is a Π-structured quasi-idempotent matrix L_4 with $d(L_4, B_{\sigma_4}) \leqslant \varepsilon'$.*

Lemma 3.3. *For each natural number n there is a natural number $p = p(n)$ and a natural number $k_0 = k_0(n)$ with the following property: Let $\mathscr{A}$ be a closed subset of $\mathscr{S}_n$. There is an integer $R = R(\mathscr{A})$ such that for any $k \geqslant k_0$, if $M \in \mathscr{A}^{(k)}$, then for any positive integer i there is a matrix $N_i \in \mathscr{A}^{(k+ip)}$ such that $N_i \leqslant RM$.*

The reader should note the similarity between Lemmas 3.1 and 3.3; the latter can be viewed as a very weak form of the former.

3.1. Proof of Lemma 3.1 from Lemmas 3.3 and 3.2

The number $p(n)$ in Lemma 3.1 is taken to be the number $p(n)$ in Lemma 3.3. In the hypothesis of Lemma 3.1 we are given $\mathscr{A}$ and ε. Choose $R = R(\mathscr{A})$ and $k_0 = k_0(n)$ as in Lemma 3.3 and define $\varepsilon' = \frac{\varepsilon}{8nR}$. Choose $b = b(n, \varepsilon')$ as given by Lemma 3.2 and define $m_0 = k_0 b$. (Note that since ε' and k_0 are determined by $\mathscr{A}$ and ε, m_0 is also determined by $\mathscr{A}$ and ε.)

We are given $m \geqslant m_0$, and $(A_1, A_2, \ldots, A_m) \in \mathscr{A}^m$ and an integer $i \geqslant 1$ and we want to find a matrix in $\mathscr{A}^{(m+ip)}$ that is within ε of the product $A_{[1,m]}$. Consider the first m_0 matrices $A_1, \ldots, A_{m_0}$ and group them into b blocks of size k_0. Define stochastic matrices $B_1, B_2, \ldots, B_b$ where B_i is the product of the k_0 matrices belonging to the ith block. Apply Lemma 3.2 to get an interval chain σ of length 5 spanning $[b]$, and define L_2, L_3, L_4 as in that lemma. Let $k = |\sigma_3| k_0$ and apply Lemma 3.3 with $M = B_{\sigma_3}$. Hence for $i \geqslant 1$, there exists a matrix $N_i \in \mathscr{A}^{(k+ip)}$ such that $N_i \leqslant R B_{\sigma_3}$. Now, since B_{σ_3} is within ε' of the Π-absorbing matrix L_3, we must have $B_{\sigma_3}(s,t) \leqslant \varepsilon'$ for any $(s,t) \notin G(\Pi)$. Therefore $N_i(s,t) \leqslant R\varepsilon' = \frac{\varepsilon}{8n}$ for any $(s,t) \notin G(\Pi)$, and we conclude from Proposition 2.9 that $M_i = \tilde{N}_i \langle G(\Pi) \rangle$ is Π-absorbing and $d(N_i, M_i) \leqslant \frac{\varepsilon}{4}$.

The matrix $C_i = B_{\sigma_1} B_{\sigma_2} N_i B_{\sigma_4} B_{\sigma_5} A_{[m_0+1,m]}$ belongs to $\mathscr{A}^{(m+ip)}$. To complete the proof it suffices to show:

Claim. $d(B_{\sigma_1} B_{\sigma_2} N_i B_{\sigma_4} B_{\sigma_5} A_{[m_0+1,m]}, B_{\sigma_1} B_{\sigma_2} B_{\sigma_3} B_{\sigma_4} B_{\sigma_5} A_{[m_0+1,m]}) \leqslant \varepsilon$.

By Proposition 2.6, the expression on the left is at most $d(B_{\sigma_2} N_i B_{\sigma_4}, B_{\sigma_2} B_{\sigma_3} B_{\sigma_4})$. Now:

$$\begin{aligned} d(B_{\sigma_2} B_{\sigma_3} B_{\sigma_4}, B_{\sigma_2} N_i B_{\sigma_4}) \leqslant{} & d(B_{\sigma_2} B_{\sigma_3} B_{\sigma_4}, L_2 L_3 L_4) + d(L_2 L_3 L_4, L_2 M_i L_4) \\ & + d(L_2 M_i L_4, B_{\sigma_2} N_i B_{\sigma_4}) \end{aligned} \tag{1}$$

by Proposition 2.6, $d(B_{\sigma_2} B_{\sigma_3} B_{\sigma_4}, L_2 L_3 L_4) \leqslant d(B_{\sigma_2}, L_2) + d(B_{\sigma_3}, L_3) + d(B_{\sigma_4}, L_4)$. Each of these three summands is at most ε', so $d(B_{\sigma_2} B_{\sigma_3} B_{\sigma_4}, L_2 L_3 L_4) \leqslant 3\varepsilon' \leqslant \frac{3\varepsilon}{8}$. Similarly, since $d(M_i, N_i) \leqslant \frac{\varepsilon}{4}$, $d(L_2 M_i L_4, B_{\sigma_2} N_i B_{\sigma_4}) \leqslant \frac{\varepsilon}{4} + 2\varepsilon' \leqslant \frac{\varepsilon}{2}$.

Since L_2 is $Res(\Pi)$-avoiding and L_3 is Π-structured quasi-idempotent and M_i and L_3 are each Π-absorbing, Lemma 2.8 implies $L_2M_iL_4 = L_2L_3L_4$. Thus $d(L_2L_3L_4, L_2M_iL_4) = 0$ and the sum on the right hand side of (1) is at most ε, proving the claim and the lemma.

So it remains to prove Lemmas 3.2 and 3.3. In Section 3.2 we present a lemma that is key to the proof of both of these lemmas. In Section 3.3 we prove Lemma 3.3. In Section 3.4 we present another lemma which is used in Section 3.5 to prove Lemma 3.2.

3.2. Partitioning a sequence of matrices into many segments with the same pattern

The following lemma asserts that given any long enough sequence of stochastic matrices, it is possible to find a stable digraph D and a chain of segments of length h such that each of the h subproducts can be very well approximated by a matrix (not the same for each subproduct) that has pattern D, where the closeness of the approximation is small relative to the smallest nonzero entry of the approximating matrix.

First we need a definition. If D is a digraph and ω, δ are real numbers in $[0, 1]$ we say that a stochastic matrix A is (D, ω, δ)-*conforming* if for all $(s, t) \in D$, $A(s, t) \geqslant \omega$ and for all $(s, t) \notin D$, $A(s, t) < \omega\delta$. (Intuitively, the entries corresponding to D are "big" and the other entries are "small".)

Lemma 3.4. *Let n, h be positive integers and $\delta > 0$. There exists a positive integer $r_0 = r_0(n, h)$ (independent of δ) and a real number $\delta' = \delta'(n, \delta)$ (independent of h) with the following property: Given any $r \geqslant r_0$ and $(A_1, A_2, \ldots, A_r) \in (\mathscr{S}_n)^r$ there exists a stable digraph D on $[n]$, a real number $\omega \in [\delta', 1]$ and an interval chain $(\sigma_1, \sigma_2, \ldots, \sigma_h)$ of length h contained in $[r_0]$ such that for each $i \in [h]$, A_{σ_i} is (D, ω, δ)-conforming.*

Proof. Let n, h, δ be given. We first define a sequence γ of $n^2 + 3$ real numbers $0 = \gamma_0 < \gamma_1 < \gamma_2 < \cdots < \gamma_{n^2+1} < \gamma_{n^2+2} = 1$ with $\gamma_{n^2+1} = 1/(n + 1)$, and for $j \in [1, n^2]$, $\gamma_j = (\gamma_{j+1}/n)^{n!}\delta$. The number δ' is defined to be $(\gamma_2)^{n!}$. Observe that δ' is independent of h.

Define the real intervals $I_0, I_1, \ldots, I_{n^2+1}$ by $I_j = (\gamma_j, \gamma_{j+1}]$. These intervals partition $(0, 1]$. For $A \in \mathscr{S}_n$, let l_A be the largest index j such that I_j contains no entry of A; j is well defined and positive by the pigeonhole principle. Also, $j \neq n^2 + 1$ since A contains an entry that is at least $1/n$. Let G_A be the graph consisting of all (s, t) such that $A(s, t) \geqslant \gamma_{l_A+1}$, i.e., $A(s, t)$ lies in one of the intervals to the right of I_{l_A}. The ordered pair (l_A, G_A) is called the *type* of the matrix A.

Now define r_0 to be $(n!h)^{n^2 2^{n^2}}$. Let $r \geqslant r_0$ and suppose that $A_1, A_2, \ldots, A_r$ is a sequence of stochastic matrices. Assign to each (integer) subinterval σ a "color" which is the type $(l_{A_\sigma}, G_{A_\sigma})$ of the matrix A_σ. There are at most $n^2 2^{n^2}$ different colors (n^2 is a bound on the number of distinct values for l_A and 2^{n^2} is the number of different digraphs). So by Lemma 2.4 and the fact that $r \geqslant r_0$, there is an interval chain τ of length $n!h$ in which each interval has the same color. Let (l, G) be the common color of these intervals.

We now specify D, ω, and σ required by the conclusion of the lemma. Let $D = G^{n!}$. By Lemma 2.5, D is a stable digraph. Let $\omega = \gamma_l^{n!}$. Define $(\sigma_1, \ldots, \sigma_h)$ as follows: group the intervals of τ consecutively into h groups of size $n!$ and let σ_i be the union of the intervals in the ith group.

Let $i \in [h]$, we must show that A_{σ_i} is (D, ω, δ)-conforming. A_{σ_i} is the product of $n!$ matrices, each of type (l, G). Thus for each of these $n!$ matrices, each entry corresponding to an arc of G is at least γ_{l+1} and each entry corresponding to a non-arc of G is less than γ_l. Since $D = G^{n!}$, each entry of A_{σ_j} corresponding to an arc of D is at least $\gamma_{l+1}^{n!} = \omega$. Each entry corresponding to a non-arc of D is the sum of at most $n^{n!}$ terms each of which is less than γ_l, and so each such entry is less than $n^{n!}\gamma_l = \delta\omega$. □

3.3. *Proof of Lemma 3.3*

We will need the following fact (which is a special case of a property of compact subsets of Euclidean space):

Lemma 3.5. *Let $\mathscr{B} \subseteq \mathscr{S}_n$ be closed. There exists a number $\alpha = \alpha(\mathscr{B}) > 0$ with the following property. Given any matrix $A \in \mathscr{B}$, there is a matrix $B \in \mathscr{B}$ such that $B(s,t) = 0$ for all (s,t) such that $A(s,t) < \alpha$.*

Proof. Suppose no such α exists. Then for each $i \geqslant 1$, the number 2^{-i} violates the property, so there is a matrix $A_i \in \mathscr{B}$ such that for any matrix $B \in \mathscr{B}$, there is an (s,t) such that $A_i(s,t) < 2^{-i}$ and $B(s,t) > 0$. Define the digraph $D_i = \{(s,t) : A_i(s,t) \geqslant 2^{-i}\}$. Choose D such that $D_i = D$ for infinitely many i and consider the subsequence of matrices A_i such that $D_i = D$; by compactness there is an infinite subsequence $i_1, i_2, i_3, \ldots$ such that $\{A_{i_j}\}$ converges to a matrix $B \in \mathscr{B}$. We now have a contradiction to the choice of A_{i_1}: By definition of D and i_1, for any (s,t) such that $A_{i_1}(s,t) < 2^{-i_1}$ we have $(s,t) \notin D_{i_1} = D$, which implies that the sequence $\{A_{i_j}(s,t)\}$ converges to 0 and hence $B(s,t) = 0$. □

We proceed with the proof of Lemma 3.3. The number k_0 in this lemma is taken to be $r_0(n, 1)$ from Lemma 3.4 and $p(n)$ is chosen to be the least common multiple of the set $\{1, 2, \ldots, k_0\}$. Define $\delta = \min\{\alpha(\mathscr{A}^{(i)}) : 1 \leqslant i \leqslant k_0\}$ where $\alpha(\mathscr{A}^{(i)})$ is as defined in Lemma 3.5. Let $R = \frac{1}{\delta'(n,\delta)}$ where δ' is as defined in Lemma 3.4.

Suppose $(A_1, A_2, \dots, A_k) \in \mathscr{A}^k$ where $k \geqslant k_0$, and let $M = A_{[1,k]}$. By Lemma 3.4 with $h = 1$, we can find a stable digraph D, a real number $\omega \geqslant \delta'$ and an interval $\sigma \subseteq [k_0]$ such that $A_\sigma(s,t) \geqslant \omega$ for $(s,t) \in D$ and $A_\sigma(s,t) < \omega\delta$ for $(s,t) \notin D$. Let $\sigma = [l_1 + 1, r_1]$ and let $q = r_1 - l_1$. Since $q \leqslant k_0$, q is a divisor of p. By Lemma 3.5 and the fact that $\alpha(\mathscr{A}^{(q)}) \geqslant \delta$ and $\omega \leqslant 1$, there is a matrix $B \in \mathscr{A}^{(q)}$ such that $B(s,t) = 0$ for all $(s,t) \notin D$. Define the matrix $C \in \mathscr{A}^{(p)}$ to be $B^{p/q}$ and let $N_i = A_{[1,l_1]}BC^iA_{[r_1+1,k]}$. Trivially $N_i \in \mathscr{A}^{(k+ip)}$. We now show that $N_i \leqslant RA_{[1,k]}$. Since D is stable and contains the pattern of B and C^i is a power of B, $BC^i(s,t) = 0$ for all $(s,t) \notin D$. For $(s,t) \in D$, $BC^i(s,t) \leqslant 1$ while $A_\sigma(s,t) \geqslant \delta'$. We conclude that $BC^i \leqslant RA_\sigma$. Since inequalities of nonnegative matrices are preserved under pre- or post-multiplication of nonnegative matrices, $N_i = A_{[1,l_1]}BC^i A_{[r_1+1,k]} \leqslant RA$.

3.4. *Convergence of products of (D,ω,δ)-conforming matrices*

In preparation for the proof of lemma 3.2 we prove a lemma that says that given a sequence of matrices of the right length, each of which is (D, ω, δ)-conforming for some stable Π-structured digraph D and δ sufficiently small, their product is close to a $Res(\Pi)$-avoiding matrix and also to a Π-structured quasi-idempotent matrix.

First, we present a lemma assuming the stronger hypothesis that each matrix has pattern D (rather than just being close to a matrix with pattern D). This lemma is similar to standard results.

Lemma 3.6. *Let Π be a partial partition of $[n]$ and D be a stable Π-structured digraph. Suppose C_1, C_2, Γ, C_m are stochastic matrices with pattern D. Let $\gamma = \mathrm{e}^{-\sum_i \mu(C_i)}$.*

1. *$C_{[1,m]}$ is within distance 2γ of some $Res(\Pi)$-avoiding matrix.*
2. *$C_{[1,m]}$ is within distance $n\gamma^2$ of some Π-structured quasi-idempotent matrix.*

Proof. Since each C_i has pattern D and D is stable, any product $C = \prod_i C_i$ has pattern D.

To prove (1), let $\Pi = \Gamma_D$ and let $S = Res(\Pi)$. and consider the matrix $\tilde{C}\langle S\rangle$. This matrix is S-absorbing and we will show $d(C, \tilde{C}\langle S\rangle) \leqslant 2\gamma$. By Proposition 2.9, it suffices to show that $C(s,S) \leqslant \gamma$ for each $s \in [n]$. For $s \in \cup\Pi$ we have $C(s,S) = 0$, since D is Π-absorbing. For $s \in S$, view C as the transition probability matrix for the m-step stochastic process defined by $(C_1, C_2, \dots, C_m)$. Note that if the process ever leaves S it never returns, so $C(s,S)$ is equal to the probability that, starting the process from s, the process is in S after every step. Since each C_i has pattern D and D is stable, $C_i(t, \cup\Pi)$ must be nonzero for each state t and therefore it is at least $\mu(C_i)$. In other words for each i and state $t \in S$, the conditional probability that the process is in S after step i given that it is in t after step $i-1$ is at most $(1 - \mu(C_i))$. We conclude that $C(s,S) \leqslant \prod_{i=1}^m (1 - \mu(C_i)) \leqslant \mathrm{e}^{-\sum_{i=1}^m \mu(C_i)}$.

Now for the proof of (2). By Propostion 2.10 it suffices to show that $\Delta_C \leqslant \gamma^2$. We show by reverse induction on i that for $i \in [m]$, $\Delta_{C_{[i,m]}} \leqslant \prod_{j=i}^{m}(1-2\mu(C_j))$, from which the desired inequality follows.

For the basis $i = m$, note that if t is recurrent and $|\Pi[t]| = 1$ then $\Delta_{C_m}(t) = 0$. Otherwise, $C_m(s,t) \geqslant \mu(C_i)$ for all $s \in \Pi[t]$ (since D is stable) and so $\max_{C_m}(t) \leqslant 1-\mu(C_i)$ and $\min_{C_m}(t) \geqslant \mu(C_i)$; thus $\Delta_{C_m}(t) \leqslant 1-2\mu_i$.

The induction step follows immediately from:

Claim. If A, B are matrices with pattern D where D is stable, then $\Delta_{AB} \leqslant (1-2\mu(A))\Delta_B$.

For the claim, let t be an arbitrary recurrent state. We first upper bound $\max_{AB}(t)$. Let $u_t \in \Pi[t]$ be the state minimizing $B(u,t)$ over all $u \in \Pi[t]$. Then for $s \in \Pi[t]$ we have,

$$\begin{aligned} AB(s,t) &= \sum_{u\in\Pi[s]} A(s,u)B(u,t) \\ &\leqslant \sum_{u\in\Pi[s]-\{u_t\}} A(s,u)\max_B(t) + A(s,u_t)\min_B(t) \\ &= \max_B(t) - A(s,u_t)\Delta_B(t) \leqslant \max_B(t) - \mu(A)\Delta_B(t). \end{aligned}$$

So $\max_{AB}(t) \leqslant \max_B(t) - \mu(A)\Delta_B(t)$. A similar argument gives $\min_{AB}(t) \geqslant \min_B(t) + \mu(A)\Delta_B(t)$. Combining these gives $\Delta_{AB}(t) \leqslant \Delta_B(t)(1-2\mu(A))$ for each recurrent t, and the claim follows. □

We now prove an "approximate" version of the previous lemma, in which the matrices are assumed only to be (D, ω, δ) conforming for some appropriate δ.

Lemma 3.7. *Let n, ε' be given, let $J = \lceil \ln \frac{4n}{\varepsilon'} \rceil$, $\delta = \frac{\varepsilon'}{8nJ}$, and $\omega \in [0, 1]$. Let Π be a partial partition and D be a stable Π-structured digraph. If $K = J\lceil 1/\omega \rceil$ then the product of any K matrices, each (D, ω, δ)-conforming, is within ε' of some $Res(\Pi)$-avoiding matrix, and is also within ε' of some Π-structured quasi-idempotent matrix.*

Proof. Let D be a stable digraph and $\omega \in [0, 1]$ and $K = J\lceil 1/\omega \rceil$, so that $J/\omega \leqslant K \leqslant 2J/\omega$. Let $A_1, A_2, \ldots, A_K$ be a sequence of (D, ω, δ)-conforming matrices and for each $i \in [K]$, let $B_i = \widetilde{A_i}\langle D \rangle$. By Proposition 2.9, $d(A_i, B_i) \leqslant 2n\omega\delta$ and by Proposition 2.6, $d(A_{[1,K]}, B_{[1,K]}) \leqslant 2n\omega\delta K \leqslant \varepsilon'/2$.

Each B_i has pattern D and smallest nonzero entry at least ω. Therefore, by Lemma 3.6, $B_{[1,K]}$ is within distance $n\mathrm{e}^{-2K\omega} \leqslant \varepsilon'/2$ of some Π-structured quasi-idempotent matrix, and so $A_{[1,K]}$ is within ε' of that matrix. Similarly, by Lemma 3.6, $B_{[1,K]}$ is within $2\mathrm{e}^{-K\omega} \leqslant \varepsilon'/2$ of some $Res(\Pi)$-avoiding matrix and hence $A_{[1,K]}$ is within ε' of that matrix. □

3.5. Proof of Lemma 3.2

We are given n and $\varepsilon' > 0$. We will define a number $b = b(n, \varepsilon')$ and show that given $(B_1, B_2, \ldots, B_b) \in \mathscr{S}^b$ we can find the desired spanning interval chain $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ of $[b]$.

Define $\delta = \delta(n, \varepsilon')$ and $J = J(n, \varepsilon')$ as in Lemma 3.7. Let $\delta' = \delta'(n, \delta)$ be defined as in Lemma 3.4. Define $h = 2J\lceil 1/\delta' \rceil + 1$. Finally, define $b = r_0(n, h)$ as in Lemma 3.4. Note that b can be expressed as a function of n and ε'.

Given $(B_1, B_2, \ldots, B_b)$, apply Lemma 3.4 to obtain an interval chain $(\tau_1, \tau_2, \ldots, \tau_h)$ contained in $[b]$, a stable digraph D and a real number $\omega \geqslant \delta'$ so that each B_{σ_i} is (D, ω, δ)-conforming. Define $K = K(J, \omega)$ as in Lemma 3.7 and note that $h \geqslant 2K + 1$.

We now define σ_1 to be the portion of $[b]$ preceding τ_1, $\sigma_2 = \cup_{i=1}^{K} \tau_i$, $\sigma_3 = \tau_{K+1}$, $\sigma_4 = \cup_{i=K+2}^{2K+1} \tau_i$ and σ_5 is the portion of $[b]$ coming after σ_4. We also define $\Pi = \Pi_D$.

By Lemma 3.7 we have that B_{σ_2} is within ε' of some $Res(\Pi)$-avoiding matrix, and B_{σ_4} is within ε' of some Π-structured quasi-idempotent matrix. Finally B_{σ_3} is (D, ω, δ)-conforming, so by Proposition 2.9, is within $2n\omega\delta \leqslant \varepsilon'$ of $\tilde{B}_{\sigma_3}$ which has pattern D and is therefore Π-absorbing.

Acknowledgements

We express our gratitude to Joseph Wong for pointing us to Ref. [4], and to D.J. Hartfiel for helpful comments.

References

[1] A. Condon, L. Hellerstein, S. Pottle, A. Wigderson, On the power of finite automata with both non-deterministic and probabilistic states, SIAM J. Comput. 27 (1998) 739–762.
[2] C. Dwork, L. Stockmeyer, Finite state verifiers I: the power of interaction, JACM 39 (4) (1992) 800–828.
[3] P. Erdös, G. Szekeres, A combinatorial problem in geometry, Compos. Math. 2 (1935) 464–470.
[4] D.J. Hartfiel, Markov Set-Chains, Springer, Berlin, 1998.
[5] D.J. Hartfiel, Nonhomogeneous Matrix Products, World Scientific, 2002.