Markov chains and applications

Dr. András Pongrácz

Week 6, University of Debrecen

Irreducibility and regularity

Terminology

Reminder : irreducibility means that the associated graph is strongly connected.

Terminology

Reminder : irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix,

Terminology

Reminder : irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix, that is, one with a positive power.

Terminology

Reminder: irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix, that is, one with a positive power. It could be called primitive, but regular is a more popular word for this property.

Reminder: irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix, that is, one with a positive power. It could be called primitive, but regular is a more popular word for this property. Also note that sometimes irreducible Markov chains are called ergodic.

Terminology

Reminder: irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix, that is, one with a positive power. It could be called primitive, but regular is a more popular word for this property. Also note that sometimes irreducible Markov chains are called ergodic.

To complicate matters, sometimes irreducible aperiodic Markov chains are called ergodic.

Terminology

Reminder: irreducibility means that the associated graph is strongly connected.

A finite Markov chain is regular if it has a primitive transition matrix, that is, one with a positive power. It could be called primitive, but regular is a more popular word for this property. Also note that sometimes irreducible Markov chains are called ergodic.

To complicate matters, sometimes irreducible aperiodic Markov chains are called ergodic. These two conditions together turn out to be equivalent to primitivity, that is, regularity.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Proposition

If A is irreducible, then it has the same period at all indices.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Proposition

If A is irreducible, then it has the same period at all indices.

The proof is an exercise.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Proposition

If A is irreducible, then it has the same period at all indices.

The proof is an exercise.

This common period at all indices is called the period p of A.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Proposition

If A is irreducible, then it has the same period at all indices.

The proof is an exercise.

This common period at all indices is called the period p of A. If $p = 1$,

we say that A is aperiodic.

Period

Definition

Given a non-negative matrix $A \in M_n(\mathbb{R})$, the period at index i is $\gcd\{k \in \mathbb{N} \mid M^k[i,i] > 0\}$.

Proposition

If A is irreducible, then it has the same period at all indices.

The proof is an exercise.

This common period at all indices is called the period p of A. If $p = 1$, we say that A is aperiodic. E.g., $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ has period $p = 3$.

Primitive matrices

Period

Theorem

A non-negative matrix A is primitive (i.e., regular) iff it is irreducible and aperiodic.

Period

Theorem

A non-negative matrix A is primitive (i.e., regular) iff it is irreducible and aperiodic.

We first prove a technical number theoretic lemma.

Lemma

Given positive integers $x_1, x_2, \ldots, x_k$ with $\gcd(x_1, x_2, \ldots, x_k) = 1$, there exists an $N \in \mathbb{N}$ such that all positive integers $M \geq N$ can be expressed as a linear combination $M = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$ with coefficients $\alpha_i \in \mathbb{N}$.

Proof

Proof of Lemma : By induction on k.

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \gcd(x_1, \ldots, x_{k-1})$.

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \gcd(x_1, \ldots, x_{k-1})$. Then by the induction hypothesis, there is an $L \geq d$ such that for any $S \geq L$, the number Sd can be expressed as a positive integer linear combination of $x_1, \ldots, x_{k-1}$.

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \gcd(x_1, \ldots, x_{k-1})$. Then by the induction hypothesis, there is an $L \geq d$ such that for any $S \geq L$, the number Sd can be expressed as a positive integer linear combination of $x_1, \ldots, x_{k-1}$.

If $d = 1$, then $N = L + x_k$ is clearly appropriate. Let $d \geq 2$.

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \mathrm{gcd}(x_1, \ldots, x_{k-1})$. Then by the induction hypothesis, there is an $L \geq d$ such that for any $S \geq L$, the number Sd can be expressed as a positive integer linear combination of $x_1, \ldots, x_{k-1}$.

If $d = 1$, then $N = L + x_k$ is clearly appropriate. Let $d \geq 2$. As $\mathrm{gcd}(x_1, \ldots, x_{k-1}, x_k) = \mathrm{gcd}(d, x_k) = 1$, there exist integers a, b such that $ad + bx_k = 1$,

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \gcd(x_1, \ldots, x_{k-1})$. Then by the induction hypothesis, there is an $L \geq d$ such that for any $S \geq L$, the number Sd can be expressed as a positive integer linear combination of $x_1, \ldots, x_{k-1}$.

If $d = 1$, then $N = L + x_k$ is clearly appropriate. Let $d \geq 2$. As $\gcd(x_1, \ldots, x_{k-1}, x_k) = \gcd(d, x_k) = 1$, there exist integers a, b such that $ad + bx_k = 1$, with $0 \leq a < x_k$ (note that $b < 0$).

Proof

Proof of Lemma: By induction on k. The initial case $k = 1$ is clear, as then $x_1 = 1$.

Assume that $k > 1$ and the assertion holds for smaller k. Let $d = \gcd(x_1, \ldots, x_{k-1})$. Then by the induction hypothesis, there is an $L \geq d$ such that for any $S \geq L$, the number Sd can be expressed as a positive integer linear combination of $x_1, \ldots, x_{k-1}$.

If $d = 1$, then $N = L + x_k$ is clearly appropriate. Let $d \geq 2$. As $\gcd(x_1, \ldots, x_{k-1}, x_k) = \gcd(d, x_k) = 1$, there exist integers a, b such that $ad + bx_k = 1$, with $0 \leq a < x_k$ (note that $b < 0$). We show that $N = Ldx_k$ is a proper choice.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then $M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then $M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

Proof of Theorem: For all $1 \leq i \leq n$, let $x_{i1}, \ldots, x_{ik_i}$ be numbers such that $A^{x_{ij}}[i, i] > 0$ and $\gcd(x_{i1}, \ldots, x_{ik_i}) = 1$.

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then $M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

Proof of Theorem: For all $1 \leq i \leq n$, let $x_{i1}, \ldots, x_{ik_i}$ be numbers such that $A^{x_{ij}}[i, i] > 0$ and $\gcd(x_{i1}, \ldots, x_{ik_i}) = 1$. Let N_i be the positive integer provided by the Lemma for $x_{i1}, \ldots, x_{ik_i}$, and let N_0 be the largest N_i.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then $M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

Proof of Theorem: For all $1 \leq i \leq n$, let $x_{i1}, \ldots, x_{ik_i}$ be numbers such that $A^{x_{ij}}[i, i] > 0$ and $\gcd(x_{i1}, \ldots, x_{ik_i}) = 1$. Let N_i be the positive integer provided by the Lemma for $x_{i1}, \ldots, x_{ik_i}$, and let N_0 be the largest N_i.

Then for all $1 \leq i \leq n$, any integer $M \geq N_0$ can be expressed as a positive integer linear combination of $x_{i1}, \ldots, x_{ik_i}$.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then $M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

Proof of Theorem: For all $1 \leq i \leq n$, let $x_{i1}, \ldots, x_{ik_i}$ be numbers such that $A^{x_{ij}}[i, i] > 0$ and $\gcd(x_{i1}, \ldots, x_{ik_i}) = 1$. Let N_i be the positive integer provided by the Lemma for $x_{i1}, \ldots, x_{ik_i}$, and let N_0 be the largest N_i.

Then for all $1 \leq i \leq n$, any integer $M \geq N_0$ can be expressed as a positive integer linear combination of $x_{i1}, \ldots, x_{ik_i}$. By irreducibility, for all i, j there is a $K(i, j) \geq 0$ such that we can reach the vertex with index j in G_A from the vertex with index i in $K(i, j)$ steps.

Proof

So let $M \geq Ldx_k$. Write M in the form $Sd - C$ with $0 \leq C < d$. In particular, $S > Lx_k > La$. Then
$M = Sd - C(ad + bx_k) = (S - Ca)d + (-b)x_k$, where $-b > 0$ and $S - Ca > (L - C)a > a$, since $L \geq d > C$.

Proof of Theorem: For all $1 \leq i \leq n$, let $x_{i1}, \ldots, x_{ik_i}$ be numbers such that $A^{x_{ij}}[i, i] > 0$ and $\gcd(x_{i1}, \ldots, x_{ik_i}) = 1$. Let N_i be the positive integer provided by the Lemma for $x_{i1}, \ldots, x_{ik_i}$, and let N_0 be the largest N_i.

Then for all $1 \leq i \leq n$, any integer $M \geq N_0$ can be expressed as a positive integer linear combination of $x_{i1}, \ldots, x_{ik_i}$. By irreducibility, for all i, j there is a $K(i, j) \geq 0$ such that we can reach the vertex with index j in G_A from the vertex with index i in $K(i, j)$ steps. Let K be the largest $K(i, j)$. Putting $N = N_0 + K$, we have that A^N is a positive matrix.

Periodic matrices

$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ has period $p = 3$.

$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ has period $p = 3$.

Its eigenvalues are the third roots of unity $1, \varepsilon, \varepsilon^2$ with corresponding eigenvectors $(1,1,1)^*, (1, \varepsilon, \varepsilon^2)^*, (1, \varepsilon^2, \varepsilon)^*$.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$ has period $p = 3$.

Its eigenvalues are the third roots of unity $1, \varepsilon, \varepsilon^2$ with corresponding eigenvectors $(1,1,1)^*, (1, \varepsilon, \varepsilon^2)^*, (1, \varepsilon^2, \varepsilon)^*$.

Note that all three eigenvalues have absolute value 1, which is impossible in the aperiodic irreducible case.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \text{ has period } p = 3.$$

Its eigenvalues are the third roots of unity $1, \varepsilon, \varepsilon^2$ with corresponding eigenvectors $(1,1,1)^*, (1, \varepsilon, \varepsilon^2)^*, (1, \varepsilon^2, \varepsilon)^*$.

Note that all three eigenvalues have absolute value 1, which is impossible in the aperiodic irreducible case.
We need to go through the proof of the Gershgorin theorem to see that there are always p roots with modulus 1, and all of them are p-th roots of unity.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$.
Let v_i be a coordinate with largest absolute value;

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$. Let v_i be a coordinate with largest absolute value; WLOG $v_i = 1$.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$.
Let v_i be a coordinate with largest absolute value ; WLOG $v_i = 1$.

Then $(A^p \underline{v})[i] = (\lambda^p \underline{v})[i] = \lambda^p v_i$.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$. Let v_i be a coordinate with largest absolute value; WLOG $v_i = 1$.

Then $(A^p \underline{v})[i] = (\lambda^p \underline{v})[i] = \lambda^p v_i$.

Taking the absolute value (cf. the proof of Gershgorin) shows that for every index j attainable in p steps from i in the digraph,

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$. Let v_i be a coordinate with largest absolute value; WLOG $v_i = 1$.

Then $(A^p\underline{v})[i] = (\lambda^p\underline{v})[i] = \lambda^p v_i$.

Taking the absolute value (cf. the proof of Gershgorin) shows that for every index j attainable in p steps from i in the digraph, we have $v_j = 1$, and then $\lambda^p = 1$.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$. Let v_i be a coordinate with largest absolute value; WLOG $v_i = 1$.

Then $(A^p \underline{v})[i] = (\lambda^p \underline{v})[i] = \lambda^p v_i$.

Taking the absolute value (cf. the proof of Gershgorin) shows that for every index j attainable in p steps from i in the digraph, we have $v_j = 1$, and then $\lambda^p = 1$. These blocks of states partition the underlying set of states into p classes.

Given an eigenvalue λ with $|\lambda| = 1$ and corresponding eigenvector $\underline{v}$. Let v_i be a coordinate with largest absolute value; WLOG $v_i = 1$.

Then $(A^p \underline{v})[i] = (\lambda^p \underline{v})[i] = \lambda^p v_i$.

Taking the absolute value (cf. the proof of Gershgorin) shows that for every index j attainable in p steps from i in the digraph, we have $v_j = 1$, and then $\lambda^p = 1$. These blocks of states partition the underlying set of states into p classes.

These observations lead to the generalization of the Perron-Frobenius theorem; see Class 7.

Exercises

Exercises

1. Prove that given an irreducible matrix A and index i, the numbers k such that $(A^k)[i,i] > 0$ form a sequence that is eventually arithmetic: that is, there is a large enough K and number p, r such that for $m > K$, an exponent is good iff it is of the form $p\ell + r$.

Exercises

1. Prove that given an irreducible matrix A and index i, the numbers k such that $(A^k)[i,i] > 0$ form a sequence that is eventually arithmetic: that is, there is a large enough K and number p, r such that for $m > K$, an exponent is good iff it is of the form $p\ell + r$.
2. By using a density argument, prove that p is the same for all indices i. Conclude that every vertex has the same period.

Exercises

1. Prove that given an irreducible matrix A and index i, the numbers k such that $(A^k)[i, i] > 0$ form a sequence that is eventually arithmetic : that is, there is a large enough K and number p, r such that for $m > K$, an exponent is good iff it is of the form $p\ell + r$.
2. By using a density argument, prove that p is the same for all indices i. Conclude that every vertex has the same period.
3. Find an irreducible, aperiodic chain such that the (square) zero-matrices along the diagonal in the "canonical" form PAP^{-1} are of different sizes, the non-zero blocks A_j are square, and the period p does not divide n.