Applying the same argument to A^T, which has the same λ_0 as A, yields the row sum bounds. □

Corollary 1.10 *Let $P \geq 0$ be the transition matrix of a regular Markov chain. Then there exists a unique distribution vector π such that $\pi P = \pi$. ($\Leftrightarrow$ $P^T \pi^T = \pi^T$)*

Proof. By Lemma 1.6 and Corollary 1.8, P has a unique largest eigenvalue $\lambda_0 \in \mathbb{R}$. By Proposition 1.9, $\lambda_0 = 1$, because as a stochastic matrix all row sums of P (i.e. the column sums of P^T) are 1. Since the geometric multiplicity of λ_0 is 1, there is a unique stochastic vector π (i.e. satisfying $\sum_i \pi_i = 1$) such that $\pi P = \pi$. □

1.3 Convergence of Regular Markov Chains

In Corollary 1.10 we established that a regular Markov chain with transition matrix P has a unique stationary distribution vector π such that $\pi P = \pi$.

By elementary arguments (page 2) we know that starting from any initial distribution q, if the iteration $q, qP, qP^2, \ldots$ converges, then it must converge to this unique stationary distribution.

However, it remains to be shown that if the Markov chain determined by P is regular, then the iteration always converges.

The following matrix decomposition is well known:

Lemma 1.11 (Jordan canonical form) *Let $A \in \mathbb{C}^{n \times n}$ be any matrix with eigenvalues $\lambda_1, \ldots, \lambda_l \in \mathbb{C}$, $l \leq n$. Then there exists an invertible matrix $U \in \mathbb{C}^{n \times n}$ such that*

$$UAU^{-1} = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_r \end{bmatrix}$$

*where each J_i is a $k_i \times k_i$ **Jordan block** associated to some eigenvalue λ of A:*

$$J_i = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

The total number of blocks associated to a given eigenvalue λ corresponds to λ's geometric multiplicity, and their total dimension $\sum_i k_i$ to λ's algebraic multiplicity.
□

Now let us consider the Jordan canonical form of a transition matrix P for a regular Markov chain. Assume for simplicity that all the eigenvalues of P are real and distinct. (The general argument is similar, but needs more complicated notation.) Then the rows of U may be taken to be left eigenvectors of the matrix P, and the Jordan canonical form reduces to the familiar eigenvalue decomposition:

$$UPU^{-1} = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

In this case one notes that in fact the columns of $U^{-1} = V$ are precisely the *right* eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. By Lemma 1.6 and Corollary 1.8, P has a unique largest eigenvalue $\lambda_1 = 1$, and the other eigenvalues may be ordered so that $1 > |\lambda_2| \geq |\lambda_2| \geq \cdots \geq |\lambda_l|$. The unique (up to normalisation) left eigenvector associated to eigenvalue 1 is the stationary distribution π, and the corresponding unique (up to normalisation) right eigenvector is $\mathbf{1} = (1, 1, \ldots, 1)$. If the first row of U is normalised to π, then the first column of V must be normalised to $\mathbf{1}$ because $UV = UU^{-1} = I$, and hence $(UV)_{11} = u_1 v_1 = \pi v_1 = 1$.

Denoting

$$\Lambda = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix},$$

we have then:

$$P^2 = (V\Lambda U)^2 = V\Lambda^2 U = V \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^2 \end{bmatrix} U,$$

and in general

$$P^t \;=\; V\Lambda^t U \;=\; V\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^t \end{bmatrix} U$$

$$\xrightarrow[t\to\infty]{} V\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} U = \begin{bmatrix} v_{11}u_1 \\ v_{12}u_1 \\ \vdots \\ v_{1n}u_1 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix}.$$

To make the situation even more transparent, represent a given initial distribution $q = q^0$ in the eigenvector basis as

$$\begin{aligned} q &= \tilde{q}_1 u_1 + \tilde{q}_2 u_2 + \cdots + \tilde{q}_n u_n \\ &= \pi + \tilde{q}_2 u_2 + \cdots + \tilde{q}_n u_n, \quad \text{where } \tilde{q}_i = \frac{q u_i^T}{||u_i||^2}. \end{aligned}$$

Then

$$qP = (\pi + \tilde{q}_2 u_2 + \cdots + \tilde{q}_n u_n)P = \pi + \tilde{q}_2\lambda_2 u_2 + \cdots + \tilde{q}_n \lambda_n u_n,$$

and generally

$$q^{(t)} = qP^t = \pi + \sum_{i=2}^{n} \tilde{q}_i \lambda_i^t u_i,$$

implying that $q^{(t)} \xrightarrow[t\to\infty]{} \pi$, and if the eigenvalues are ordered as assumed, then

$$||q^{(t)} - \pi|| = o\left(|\lambda_2|^t\right).$$

1.4 Transient Behaviour of General Chains

So what happens to the transient states in a reducible Markov chain?

A moment's thought shows that the transition matrix of an arbitrary (finite) Markov chain can be put in the following *canonical form*:

$$P = \begin{bmatrix} P_1 & 0 & & \\ & \ddots & & 0 \\ 0 & & P_r & \\ & R & & Q \\ & & & \end{bmatrix}$$

where the r square matrices $P_1, \ldots, P_r$ in the upper left corner represent the transitions within the r minimal closed classes, Q represents the transitions among transient states, and R represents the transitions from transient states to one of the closed classes.

In this ordering, stationary distributions (left eigenvectors of P correspondonding to eigenvalue 1) must apparently be of the form $\pi = [\pi_1 \;\cdots\; \pi_r \quad 0 \;\cdots\; 0]$. (This follows e.g. from the fact that Q must be "substochastic", i.e. have at least one row sum less than 1.)

Consider then the *fundamental matrix* $M = (I-Q)^{-1}$ of the chain. Intuitively, if M is well-defined, it corresponds to $M = I + Q + Q^2 + \ldots$, and represents all the possible transition sequences the chain can have without exiting Q.

Theorem 1.12 *For any finite Markov chain, the fundamental matrix* $M = (I - Q)^{-1}$ *is well-defined and positive. Its elements can be computed from the converging series* $M = I + Q + Q^2 + \ldots$

Proof. The result will follow from some more general results to be proved later. (We will look into applications first.) □

Let i, j be any two transient states in a Markov chain with a transition matrix as above. Then:

$$\Pr(X_t = j \mid X_0 = i) = Q^t_{ij} \triangleq q^{(t)}_{ij}.$$

Thus,

$$\begin{aligned} E[\text{number of visits to } j \in T \mid X_0 = i \in T] &= q^{(0)}_{ij} + q^{(1)}_{ij} + q^{(2)}_{ij} + \ldots \\ &= I_{ij} + Q_{ij} + Q^2_{ij} + \ldots \\ &= M_{ij} \triangleq m_{ij} \end{aligned}$$

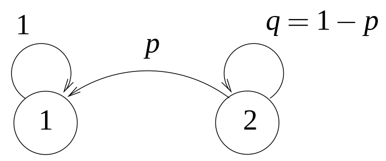

Figure 6: A Markov chain representing the geometric distribution.

Furthermore,

$$\begin{aligned} & E[\text{number of moves in } T \text{ before exiting to } C \mid X_0 = i \in T] \\ = & \sum_{j \in T} E[\text{number of visits to } j \in T \mid X_0 = i \in T] \\ = & \sum_{j \in T} m_{ij} \\ = & (M\mathbf{1})_i. \end{aligned}$$

Finally, let b_{ij} be the probability that the chain when started in transient state $i \in T$ will enter a minimal closed class via state $j \in C$. Denote $B = (b_{ij})_{i \in T, j \in C}$. Then $B = MR$.

Proof. For given $i \in T, j \in C$,

$$b_{ij} = p_{ij} + \sum_{t \in T} p_{ik} b_{kj}.$$

Thus,

$$B = R + QB \quad \Rightarrow \quad B = (I - Q)^{-1} R = MR.$$

Example 1.4 *The geometric distribution.* Consider the chain of Figure 6, arising e.g. from biased coin-flipping The transition matrix in this case is

$$P = \begin{bmatrix} 1 & 0 \\ p & q \end{bmatrix}.$$

$Q = (q)$, $M = (1-q)^{-1} = 1/p$. Thus, e.g.

$$E[\text{number of visits to 2 before exiting to 1} \mid X_0 = 2] = M\mathbf{1} = \frac{1}{p}.$$

The elementary way to obtain the same result:

$$\begin{aligned} E[\text{number of visits}] &= \sum_{t \geq 0} \Pr[\text{number of visits} \geq k] \\ &= 1 + q + q^2 + \cdots = \frac{1}{1-q} = \frac{1}{p}. \end{aligned}$$

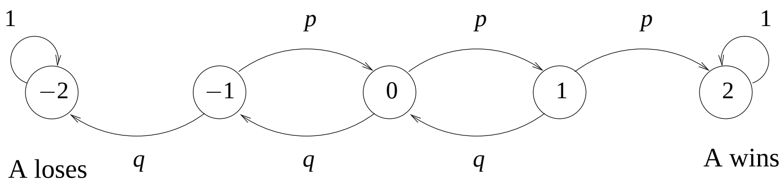

Figure 7: A Markov chain representing a coin-flipping game.

Example 1.5 *Gambling tournament.* Players A and B toss a biased coin with A's success probability equal to p and B's success probability equal to $1-p=q$. The person to first obtain n successes over the other wins. What are A's chances of winning, given that he initially has k successes over B, $-n \le k \le n$? (A more technical term for this process is "one-dimensional random walk with two absorbing barriers.")

For simplicity, let us consider only the case $n=2$. Then the chain is as represented in Figure 7, with transition matrix:

	-2	-1	0	1	2
-2	1	0	0	0	0
-1	q	0	p	0	0
0	0	q	0	p	0
1	0	0	q	0	p
2	0	0	0	0	1

i.e. in canonical form:

	-2	2	-1	0	1
-2	1	0	0	0	0
2	0	1	0	0	0
-1	q	0	0	p	0
0	0	0	q	0	p
1	0	p	0	q	0

Thus, $M=(I-Q)^{-1}$

$$= \begin{bmatrix} 1 & -p & 0 \\ -q & 1 & -p \\ 0 & -q & 1 \end{bmatrix}^{-1} = \frac{1}{p^2+q^2} \begin{bmatrix} p+q^2 & p & p^2 \\ q & 1 & p \\ q^2 & q & q+p^2 \end{bmatrix}$$

and so $B=MR$

$$= \frac{1}{p^2+q^2} \begin{bmatrix} p+q^2 & p & p^2 \\ q & 1 & p \\ q^2 & q & q+p^2 \end{bmatrix} \begin{bmatrix} q & 0 \\ 0 & 0 \\ 0 & p \end{bmatrix} = \frac{1}{p^2+q^2} \begin{bmatrix} qp+q^3 & p^3 \\ q^2 & p^2 \\ \underbrace{q^3}_{\text{A loses}} & \underbrace{pq+p^3}_{\text{A wins}} \end{bmatrix}.$$

We conclude this section by establishing the truth of Theorem 1.12 via two basic lemmas.

Lemma 1.13 *If all eigenvalues* λ *of matrix* A *satisfy* $|\lambda| < 1$*, then* $(I-A)^{-1}$ *is well-defined and satisfies*

$$(I-A)^{-1} = I + A + A^2 + \dots \tag{4}$$

Proof. Assume first that the series in (4) converges to a matrix B. Then

$$(I-A)B = (I + A + A^2 + \dots) - (A + A^2 + \dots) = I.$$

Consider first the case where A may be fully diagonalised: $M^{-1}AM = \Lambda$.

Then $A^t = M\Lambda^t M^{-1}$, and the series (4) is made up of various geometric series of the form λ_i^t, where λ_i are the eigenvalues of A. All these converge, because $|\lambda_i| < 1$. If A is not diagonalisable, there may be series of the form $\lambda_i^t, t\lambda_i^t, t^2\lambda_i^t, \dots, t^{n-1}\lambda_i^t$. Again these converge. □

Lemma 1.14 *Let* A *be a nonnegative matrix with Perron-Frobenius eigenvalue* λ_0*. Then the matrix* $(\lambda I - A)^{-1}$ *is well-defined and positive if and only if* $\lambda > \lambda_0$*.*

Proof. Suppose first that $\lambda > \lambda_0$ (≥ 0). Then the matrix $\bar{A} = A/\lambda$ has all eigenvalues less than 1 in absolute value. By Lemma 1.13:

$$(\lambda I - A)^{-1} = \frac{1}{\lambda}(I - \bar{A})^{-1} = \frac{1}{\lambda}\left(I + \frac{A}{\lambda} + \frac{A^2}{\lambda^2} + \dots\right).$$

Thus $(\lambda I - A)^{-1}$ exists and is positive since every term in the series expansion is nonnegative.

Conversely, suppose $\lambda \leq \lambda_0$. Let $x_0 \geq 0$ be an eigenvector corresponding to λ_0. Then $Ax_0 \geq \lambda x_0$, i.e. $(\lambda I - A)x_0 + p = 0$ for some $p \geq 0$. If $(\lambda I - A)^{-1}$ exists, then $(\lambda I - A)^{-1}p = -x_0$. Thus, since $p \geq 0$, $(\lambda I - A)^{-1}$ cannot be positive. □

Proof of Theorem 1.12: Since elements of Q^t are t-step transition probabilities within the transient classes, it follows (more of less from the definition of transient) that $Q^t \to 0$ as $t \to \infty$. Thus, the dominant eigenvalue of Q must be less than 1, and the claim follows from Lemma 1.14. □

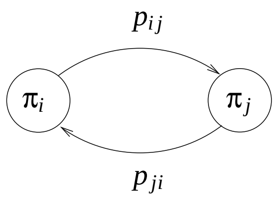

Figure 8: Detailed balance condition $\pi_i p_{ij} = \pi_j p_{ji}$.

1.5 Reversible Markov Chains

We now introduce an important special class of Markov chains often encountered in algorithmic applications. Many examples of these types of chains will be encountered later.

Intuitively, a "reversible" chain has no preferred time direction at equilibrium, i.e. any given sequence of states is equally likely to occur in forward as in backward order.

A Markov chain determined by the transition matrix $P = (p_{ij})_{i,j \in S}$ is *reversible* if there is a distribution π that satisfies the *detailed balance* conditions:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall\, i, j \in S.$$

Theorem 1.15 *A distribution satisfying the detailed balance conditions is stationary.*

Proof. It suffices to show that, assuming the detailed balance conditions, the following stationarity condition holds for all $i \in S$:

$$\pi_i = \sum_{j \in S} \pi_j p_{ji}.$$

But this is straightforward:

$$\sum_{j \in S} \pi_j p_{ji} = \sum_{j \in S} \pi_i p_{ij} = \pi_i \sum_{j \in S} p_{ji} = \pi_i.$$

□

Observe the intuition underlying the detailed balance condition: At stationarity, an equal amount of probability mass flows in each step from i to j as from j to i.(The "ergodic flows"' between states are in pairwise balance; cf. Figure 8.)

Example 1.6 *Random walks on graphs.*

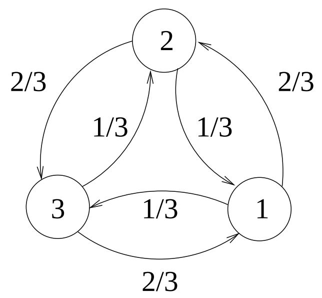

Figure 9: A nonreversible Markov chain.

Let $G = (V, E)$ be a (finite) graph, $V = \{1, \dots, n\}$. Define a Markov chain on the nodes of G so that at each step, one of the current node's neigbours is selected as the next state, uniformly at random. That is,

$$p_{ij} = \begin{cases} \frac{1}{d_i}, & \text{if } (i, j) \in E \\ 0, & \text{otherwise} \end{cases} \qquad (d_i = \deg(i))$$

Let us check that this chain is reversible, with stationary distribution

$$\pi = \begin{bmatrix} \frac{d_1}{d} & \frac{d_2}{d} & \cdots & \frac{d_n}{d} \end{bmatrix},$$

where $d = \sum_{i=1}^{n} d_i = 2|E|$. The detailed balance condition is easy to verify:

$$\pi_i p_{ij} = \begin{cases} \frac{d_i}{d} \cdot \frac{1}{d_i} = \frac{1}{d} = \frac{d_j}{d} \cdot \frac{1}{d_j} = \pi_j p_{ji}, & \text{if } (i, j) \in E \\ 0 = \pi_j p_{ji}, & \text{if } (i, j) \notin E \end{cases}$$

Example 1.7 *A nonreversible chain.*

Consider the three-state Markov chain shown in Figure 9. It is easy to verify that this chain has the unique stationary distribution $\pi = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$. However, for any $i = 1, 2, 3$:

$$\pi_i p_{i,(i+1)} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} > \pi_{i+1} p_{(i+1),i} = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}.$$

Thus, even in a stationary situation, the chain has a "preference" of moving in the counter-clockwise direction, i.e. it is not time-symmetric.