Chapter 3

Capacity

In this chapter we introduce and study the notion of capacity, which is one of the most important parameters related to constrained systems. In the context of coding, the significance of capacity will be made apparent in Chapter 4, where we show that it sets an (attainable) upper bound on the rate of any finite-state encoder for a given constrained system.

The definition of capacity is given in Section 3.1. We then provide two other characterizations of capacity—an algebraic and a probabilistic one. The algebraic characterization leads to a method for computing capacity from any lossless graph presentation of the constrained system (see Theorem 3.4 below).

3.1 Combinatorial characterization of capacity

Let S be a constrained system over an alphabet Σ and denote by $N(\ell; S)$ the number of words of length ℓ in S. The *base-2 Shannon capacity*, or simply *capacity* of S, is defined by

$$\mathsf{cap}(S) = \limsup_{\ell \to \infty} \frac{1}{\ell} \log N(\ell; S) \ .$$

Hereafter, if the base of the logarithms is omitted then it is assumed to be 2.

The Shannon capacity $\mathsf{cap}(S)$ of S measures the growth rate of the number of words of length ℓ in S, in the sense that the $N(\ell; S)$ is well-approximated by $2^{\ell\, \mathsf{cap}(S)}$ for large enough ℓ.

If $|S| = \infty$ then $0 \le \mathsf{cap}(S) \le \log |\Sigma|$. Otherwise, if S is finite, then $\mathsf{cap}(S) = -\infty$. In the latter case, there are no cycles in any presentation G of S; so each irreducible component of G is a trivial graph with one state and no edges.

Example 3.1 Let S be the $(0, 1)$-RLL constrained system which is presented by the graph G shown in Figure 3.1. For $u \in \{0,1\}$, denote by $x_u(\ell)$ the number of words of length

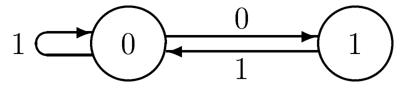

Figure 3.1: Shannon cover of the $(0,1)$-RLL constrained system.

ℓ that can be generated from state u in G. Then for $\ell \geq 1$,

$$\begin{aligned} x_0(\ell) &= x_0(\ell-1) + x_1(\ell-1) , \\ x_1(\ell) &= x_0(\ell-1) , \end{aligned}$$

and the initial conditions are obtained for $\ell = 0$ (the empty word) by

$$x_0(0) = x_1(0) = 1 .$$

So, for $\ell \geq 2$,

$$x_0(\ell) = x_0(\ell-1) + x_0(\ell-2)$$

with the initial conditions $x_0(0) = 1$, $x_0(1) = 2$. Hence, $x_0(\ell)$ are Fibonacci numbers and can be written as

$$x_0(\ell) = c_1 \lambda^\ell + c_2(-\lambda)^{-\ell} ,$$

where $\lambda = (1+\sqrt{5})/2$ (the golden mean ratio) and $c_1 > 0$. Since $\mathcal{F}_G(1) \subseteq \mathcal{F}_G(0)$ we have $\mathcal{F}_G(0) = S$ and

$$N(\ell; S) = x_0(\ell) .$$

Therefore,

$$\mathsf{cap}(S) = \log \frac{1+\sqrt{5}}{2} \approx .6942 .$$

$\square$

The 'lim sup' in the definition of capacity can be replaced by a proper limit, and this can be shown in two ways. The most direct method is as follows: first, one shows that $\log N(\ell; S)$ is a *subadditive function*—i.e., for all ℓ and m, $(\log N(\ell+m; S)) \leq (\log N(\ell; S)) + (\log N(m; S))$; then one shows that for any subadditive function $f(\ell)$, the limit $\lim_{\ell\to\infty}(f(\ell)/\ell)$ exists (see [LM95, Lemma 4.1.7]). An alternative method is provided by Theorem 3.4 below.

The following theorem shows that the capacity of a constrained system S is determined by the irreducible components of a graph presentation of S.

Theorem 3.1 *Let $S = S(G)$ be a constrained system and let $G_1, G_2, \ldots, G_k$ be the irreducible components of G, presenting the irreducible systems $S_i = S(G_i)$. Then,*

$$\mathsf{cap}(S) = \max_{1 \leq i \leq k} \mathsf{cap}(S_i) .$$

Proof. Clearly, $\mathsf{cap}(S) \geq \mathsf{cap}(S_i)$ for all $i = 1, 2, \ldots, k$. We now prove the inequality in the other direction. Any word $\mathbf{w} \in S \cap \Sigma^\ell$ can be decomposed into sub-words in the form

$$\mathbf{w} = \mathbf{w}_1 z_1 \mathbf{w}_2 z_2 \ldots z_{r-1} \mathbf{w}_r \ ,$$

where each $\mathbf{w}_j$ (possibly the empty word) is generated wholly within one of the irreducible components of G and each z_j is a label of an edge that links two irreducible components. Due to the partial ordering on the irreducible components of G, once we leave such a component we will not visit it again in the course of generating $\mathbf{w}$. Hence, $r \leq k$ and

$$N(\ell; S) = |S \cap \Sigma^\ell| \leq 2^k \cdot |\Sigma|^{k-1} \cdot \sum_{(\ell_1, \ell_2, \ldots, \ell_k)} \prod_{i=1}^{k} N(\ell_i; S_i) \ , \tag{3.1}$$

where $(\ell_1, \ell_2, \ldots, \ell_k)$ ranges over all nonnegative integer k-tuples such that $\ell_1 + \ell_2 + \ldots + \ell_k \leq \ell$. In (3.1), the term 2^k stands for the number of combinations of the traversed irreducible components; the term $|\Sigma|^{k-1}$ bounds from above the number of possible linking symbols z_j; and ℓ_i stands for the length of the sub-word of $\mathbf{w} \in S \cap \Sigma^\ell$ that is generated in the irreducible component G_i.

Without loss of generality we assume that $\mathsf{cap}(S_i)$ is nonincreasing with i and that h is the largest index i, if any, for which $\mathsf{cap}(S_i) \geq 0$; namely, $G_{h+1}, G_{h+2}, \ldots, G_k$ are the irreducible components of G with one state and no edges. Note that when no such h exists then $\mathsf{cap}(S) = -\infty$ and the theorem holds trivially.

By the definition of capacity, it follows that for every $i \leq h$ and $m \in \mathbb{N}$ we have

$$N(m; S_i) \leq \exp\Big\{m(\mathsf{cap}(S_i) + \varepsilon(m))\Big\} \leq \exp\Big\{m(\mathsf{cap}(S_1) + \varepsilon(m))\Big\} \ ,$$

where exponents are taken to base 2 and $\lim_{m \to \infty} \varepsilon(m) = 0$. Plugging this with $m = \ell_i$ into (3.1) we obtain

$$\begin{aligned} N(\ell; S) \ &\leq \ (2|\Sigma|)^k \cdot \sum_{(\ell_1, \ell_2, \ldots, \ell_h)} \exp\Big\{\Big(\sum_{i=1}^{h} \ell_i\Big) \mathsf{cap}(S_1)\Big\} \cdot \exp\Big\{\sum_{i=1}^{h} \ell_i \varepsilon(\ell_i)\Big\} \\ &\leq \ (2|\Sigma|)^k \cdot (\ell+1)^h \cdot \exp\Big\{\ell \, \mathsf{cap}(S_1)\Big\} \cdot \left(\max_{(\ell_1, \ell_2, \ldots, \ell_h)} \exp\Big\{\sum_{i=1}^{h} \ell_i \varepsilon(\ell_i)\Big\}\right) \ , \end{aligned}$$

where $(\ell_1, \ell_2, \ldots, \ell_h)$ ranges over all nonnegative integer h-tuples such that $\ell_1 + \ell_2 + \ldots + \ell_h \leq \ell$. Defining

$$\delta(\ell) = \frac{1}{\ell} \max_{(\ell_1, \ell_2, \ldots, \ell_h)} \sum_{i=1}^{h} \ell_i \varepsilon(\ell_i) \ ,$$

we obtain

$$\frac{1}{\ell} \log N(\ell; S) \leq \mathsf{cap}(S_1) + \frac{k \log(2|\Sigma|(\ell+1))}{\ell} + \delta(\ell) \ .$$

Hence, in order to complete the proof, it suffices to show that $\lim_{\ell\to\infty} \delta(\ell) = 0$. We leave this as an exercise (Problem 3.2). $\square$

The following useful fact is a straightforward consequence of the definition of capacity. The proof is left as an exercise (see Problem 3.1).

Proposition 3.2 *For any constrained system S and positive integer ℓ,*

$$\mathsf{cap}(S^\ell) = \ell \cdot \mathsf{cap}(S) .$$

3.2 Algebraic characterization of capacity

In this section, we present an algebraic method for computing the capacity of a given constrained system. This method is based on Perron-Frobenius theory of nonnegative matrices. Our full treatment of Perron-Frobenius theory is deferred to Section 3.3. Still, we will provide here a simplified version of the theorem so that we can demonstrate how it is applied to the computation of capacity. We start with the following definition.

A nonnegative real square matrix A is called *irreducible* if for every row index u and column index v there exists a nonnegative integer $\ell_{u,v}$ such that $(A^{\ell_{u,v}})_{u,v} > 0$.

For a square real matrix A, we denote by $\lambda(A)$ the *spectral radius* of A—i.e., the largest of the absolute values of the eigenvalues of A.

The following is a short version of Perron-Frobenius theorem for irreducible matrices.

Theorem 3.3 *Let A be an irreducible matrix. Then $\lambda(A)$ is an eigenvalue of A and there are right and left eigenvectors associated with $\lambda(A)$ that are strictly positive; that is, each of their components is strictly positive.*

The following theorem expresses the capacity of an irreducible constrained system S in terms of the adjacency matrix of a lossless presentation of S.

Theorem 3.4 *Let S be an irreducible constrained system and let G be an irreducible lossless (in particular, deterministic) presentation of S. Then,*

$$\mathsf{cap}(S) = \log \lambda(A_G) .$$

We break the proof of Theorem 3.4 into two lemmas.

Lemma 3.5 *Let A be an irreducible matrix. Then, for every row index u,*

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log \Big(\sum_v (A^\ell)_{u,v} \Big) = \log \lambda(A) \; .$$

Furthermore,

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log \Big(\sum_{u,v} (A^\ell)_{u,v} \Big) = \log \lambda(A) \; .$$

Proof. We make use of a positive right eigenvector $\mathbf{x}$ associated with the eigenvalue $\lambda = \lambda(A)$. Let $x_{\max}$ and $x_{\min}$ denote the maximal and minimal components of $\mathbf{x}$, respectively. Both $x_{\max}$ and $x_{\min}$ are strictly positive. For each row index u we have

$$x_{\min} \sum_v (A^\ell)_{u,v} \le \sum_v (A^\ell)_{u,v} x_v = \lambda^\ell x_u \; .$$

Thus,

$$\sum_v (A^\ell)_{u,v} \le \frac{x_u}{x_{\min}} \cdot \lambda^\ell \; .$$

Replacing $x_{\min}$ by $x_{\max}$ and reversing the direction of the inequalities, we get

$$\sum_v (A^\ell)_{u,v} \ge \frac{x_u}{x_{\max}} \cdot \lambda^\ell \; .$$

It thus follows that the ratio of $\sum_v (A^\ell)_{u,v}$ to λ^ℓ is bounded above and below by positive constants and, so, these two quantities grow at the same rate. The same holds with respect to $\sum_{u,v} (A^\ell)_{u,v}$. □

Lemma 3.6 *Let S be an irreducible constrained system and let G be an irreducible lossless presentation of S. Then,*

$$\mathsf{cap}(S) = \lim_{\ell\to\infty} \frac{1}{\ell} \log \Big(\sum_{u,v} (A_G^\ell)_{u,v} \Big) \; .$$

Proof. Recall that $\sum_{u,v} (A_G^\ell)_{u,v}$ is the number of paths of length ℓ in G. Now, every word of length ℓ in S can be generated by at least one path in G. On the other hand, since G is lossless, every word in S can be generated by at most $|V_G|^2$ paths in G. Hence, the number, $N(\ell; S)$, of words of length ℓ in S is bounded from below and above by

$$\frac{1}{|V_G|^2} \cdot \sum_{u,v} (A_G^\ell)_{u,v} \le N(\ell; S) \le \sum_{u,v} (A_G^\ell)_{u,v} \; .$$

Therefore,

$$\mathsf{cap}(S) = \lim_{\ell\to\infty} \frac{1}{\ell} \log N(\ell; S) = \lim_{\ell\to\infty} \frac{1}{\ell} \log \Big(\sum_v (A^\ell)_{u,v} \Big) = \log \lambda(A) \; .$$

Note that we have established here that 'lim sup' can indeed be replaced by a proper limit. □

Example 3.2 For the $(0, 1)$-RLL constrained system presented by the deterministic graph in Figure 3.1, the adjacency matrix is

$$A_G = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

with largest eigenvalue $\lambda = (1+\sqrt{5})/2$ and capacity $\log \lambda \approx .6942$. $\square$

Example 3.3 For $0 \leq d \leq k < \infty$, let $G(d, k)$ denote the Shannon cover in Figure 1.3 of the (d, k)-RLL constrained system. It can be shown that $\lambda(A_{G(d,k)})$ is the largest positive solution of the equation

$$z^{k+2} - z^{k+1} - z^{k-d+1} + 1 = 0$$

(see Problem 3.19). This in turn, allows to compute the capacity of any (d, k)-RLL constrained system. Table 3.1 (taken from [Imm91]) contains the capacity values of several (d, k)-RLL constrained systems. $\square$

$k \backslash d$	0	1	2	3	4	5	6
1	.6942						
2	.8791	.4057					
3	.9468	.5515	.2878				
4	.9752	.6174	.4057	.2232			
5	.9881	.6509	.4650	.3218	.1823		
6	.9942	.6690	.4979	.3746	.2669	.1542	
7	.9971	.6793	.5174	.4057	.3142	.2281	.1335
8	.9986	.6853	.5293	.4251	.3432	.2709	.1993
9	.9993	.6888	.5369	.4376	.3630	.2979	.2382
10	.9996	.6909	.5418	.4460	.3746	.3158	.2633
11	.9998	.6922	.5450	.4516	.3833	.3282	.2804
12	.9999	.6930	.5471	.4555	.3894	.3369	.2924
13	.9999	.6935	.5485	.4583	.3937	.3432	.3011
14	.9999	.6938	.5495	.4602	.3968	.3478	.3074
15	.9999	.6939	.5501	.4615	.3991	.3513	.3122
∞	1.0000	.6942	.5515	.4650	.4057	.3620	.3282

Table 3.1: Capacity values of several (d, k)-RLL constrained systems.

Example 3.4 Consider the 2-charge constrained system whose Shannon cover is given by the graph G in Figure 3.2. The adjacency matrix of G is given by

$$A_G = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

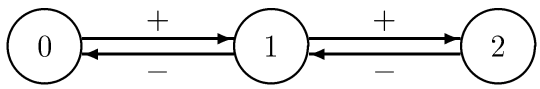

Figure 3.2: Shannon cover of the 2-charge constrained system.

with eigenvalues $\pm\sqrt{2}$ and 0. Hence, the capacity of the 2-charge constrained system is $\log\sqrt{2} = 1/2$.

More generally, if G_B is the Shannon cover in Figure 1.14 of the B-charge constrained system, then it can be shown that

$$\lambda(A_{G_B}) = 2\cos\left(\frac{\pi}{B+2}\right)$$

(see Problem 3.20). Table 3.2 lists the values of $\log\lambda(A_{G_B})$ for several values of B. $\square$

B	1	2	3	4	5	6	7	8	9	10	11	12
	.0000	.5000	.6942	.7925	.8495	.8858	.9103	.9276	.9403	.9500	.9575	.9634

Table 3.2: Capacity values of several B-charge constrained systems.

It turns out that Theorem 3.4 and Lemma 3.6 hold for any constrained system S and lossless graph G, irreducible or reducible. We show this next.

Theorem 3.7 *Let S be a constrained system and let G be a lossless presentation of S. Then, there is an irreducible constrained system $S' \subseteq S$ such that*

$$\mathsf{cap}(S') = \mathsf{cap}(S) = \log\lambda(A_G)\ .$$

Proof. Let $G_1, G_2, \ldots, G_k$ be the irreducible components of G and denote by A_i the adjacency matrix of G_i. By reordering the states, we can assume that the adjacency matrix A of G has the block-triangular form of Figure 3.3. Since the set of eigenvalues of A is the union of the set of eigenvalues of the matrices A_i, we obtain

$$\lambda(A) = \max_{i=1}^{k}\lambda(A_i)\ .$$

The result now follows from Theorem 3.1. $\square$

$$A = \begin{pmatrix} A_1 & B_{1,2} & B_{1,3} & \cdots & B_{1,k} \\ & A_2 & B_{2,3} & \cdots & B_{2,k} \\ & & A_3 & \ddots & \vdots \\ & & & \ddots & B_{k-1,k} \\ & & & & A_k \end{pmatrix} .$$

Figure 3.3: Block-triangular form.

3.3 Perron-Frobenius theory

In this section, we present a more extensive treatment of Perron-Frobenius theory. We have already exhibited one application of this theory—namely, providing a means for computing capacity. In fact, as we show in Chapters 5 and 7, this theory also serves as a major tool for constructing and analyzing constrained systems.

3.3.1 Irreducible matrices

Recall that a nonnegative real square matrix A is called irreducible if for every row index u and column index v there exists a nonnegative integer $\ell_{u,v}$ such that $(A^{\ell_{u,v}})_{u,v} > 0$. A nonnegative real square matrix that is not irreducible is called *reducible* .

The 1×1 matrix $A = \begin{pmatrix} 0 \end{pmatrix}$ will be referred to as the *trivial* irreducible matrix. The trivial irreducible matrix is the adjacency matrix of the trivial irreducible graph (which has one state and no edges).

Irreducibility of a nonnegative real square matrix A depends on the locations (row and column indexes) of the nonzero entries in A, and not on their specific values. For example, irreducibility would be preserved if we changed each nonzero entry in A to 1. Therefore, the following definition is useful.

Let A be a nonnegative real square matrix. The *support graph* of A is a graph G with a state for each row in A and an edge $u \to v$ if and only if $(A)_{u,v} > 0$. Note that A is irreducible if and only if its support graph G is irreducible, and G is irreducible if and only if its adjacency matrix A_G is irreducible.

In analogy with graphs, we can now define an *irreducible component* of a nonnegative real square matrix A as an irreducible submatrix of A whose support graph is an irreducible component of the support graph of A. The term *irreducible sink* extends to matrices in a straightforward manner. By applying the same permutation on both the rows and columns of A, we can obtain a matrix in upper block-triangular form with its irreducible components, $A_1, A_2, \ldots, A_k$, as the block diagonals, as shown in Figure 3.3.

We will use in the sequel the following notations. Let A and B be real matrices (in particular, vectors) of the same order. We write $A \geq B$ (respectively, $A > B$) if the weak (respectively, strict) inequality holds component by component. We say that A is *strictly positive* if $A > 0$.

3.3.2 Primitivity and periodicity

Let G be a nontrivial irreducible graph. We say that G is *primitive* if there exists a (strictly) positive integer ℓ such that for every ordered pair of states (u, v) of G there is a path of length ℓ from u to v. Equivalently, A_G^ℓ is strictly positive; note that this implies that A_G^m is strictly positive for every $m > \ell$, since the adjacency matrix of a nontrivial irreducible graph cannot have all-zero rows or columns. Observe that the trivial matrix is not a primitive matrix.

Let G be a nontrivial irreducible graph. The *period* of G is the greatest common divisor of the lengths of all cycles in G. We say that G is *aperiodic* if its period is 1.

It is not difficult to check that the graph in Figure 3.1 (which presents the $(0, 1)$-RLL constrained system) is aperiodic and the graph in Figure 3.2 (which presents the 2-charge constrained system) has period 2.

Proposition 3.8 *A nontrivial irreducible graph is aperiodic if and only if it is primitive.*

Proof. Let G be a primitive graph. Then there exists a positive integer ℓ such that A_G^ℓ is strictly positive, and therefore so is $A_G^{\ell+1}$. In particular, there exist cycles in G of lengths ℓ and $\ell+1$. Hence, G is aperiodic.

Conversely, assume that $G = (V, E, L)$ is aperiodic and let $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$ be cycles in G of lengths $t_1, t_2, \ldots, t_k$, respectively, such that $\gcd(t_1, t_2, \ldots, t_k) = 1$. By the extended Euclidean algorithm, there exist integers $b_1, b_2, \ldots, b_k$ such that $\sum_{i=1}^k b_i t_i = \gcd(t_1, t_2, \ldots, t_k) = 1$. Define the constants

$$M = (2|V| - 1) \max_{i=1}^{k} |b_i|$$

and

$$a_{i,j} = M - j b_i \ , \quad i = 1, 2, \ldots, k \ , \quad j = 0, 1, \ldots, 2|V| - 2 \ .$$

Note that each $a_{i,j}$ is a positive integer.

For $i = 1, 2, \ldots, k$, let u_i be the initial (and terminal) state of the cycle Γ_i and let π_i be a path from u_i to u_{i+1} in G (see Figure 3.4). For $j = 0, 1, \ldots, 2|V| - 2$, define the path γ_j by

$$\gamma_j = \Gamma_1^{a_{1,j}} \pi_1 \Gamma_2^{a_{2,j}} \pi_2 \ldots \pi_{k-1} \Gamma_k^{a_{k,j}} \ .$$

That is, γ_j starts at state u_1, then circles $a_{1,j}$ times along Γ_1, then follows the edges of π_1 to reach u_2, next circles $a_{2,j}$ times along Γ_2, and so on, until it terminates in u_k. Denote by r

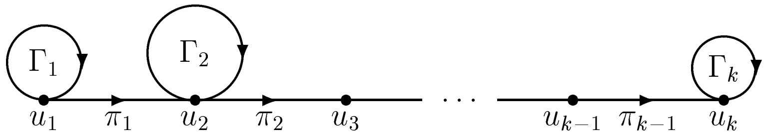

Figure 3.4: Paths for the proof of Proposition 3.8.

the length of the path $\pi_1\pi_2\ldots\pi_{k-1}$. For each $j = 0, 1, \ldots, 2|V|-2$, the length of γ_j is given by

$$r + \sum_{i=1}^{k} a_{i,j} t_i = r + \sum_{i=1}^{k} (M - jb_i)t_i = \Big(r + M \cdot \sum_{i=1}^{k} t_i\Big) - j\sum_{i=1}^{k} b_i t_i = \ell - j ,$$

where $\ell = r + M \cdot \sum_{i=1}^{k} t_i$ is independent of j.

Now, let u and v be states in G and let π_0 and π_k be the shortest paths in G from u to u_1 and from u_k to v, respectively. Since π_0 and π_k each has length smaller than $|V|$, there exists one path γ_j such that the length of $\pi_0\gamma_j\pi_k$ is exactly ℓ. Hence, $(A_G^{\ell})_{u,v} > 0$. $\square$

Lemma 3.9 *Let u and v be two states in a nontrivial irreducible graph G with period p. Then all paths in G from state u to state v have congruent lengths modulo p.*

Proof. Let γ_1 and γ_2 be two paths from u to v in G of lengths ℓ_1 and ℓ_2, respectively. Also, let γ_3 be a path of length ℓ_3 from v to u. Since $\gamma_1\gamma_3$ and $\gamma_2\gamma_3$ are cycles, their lengths must be divisible by p. Therefore,

$$\ell_1 + \ell_3 \equiv \ell_2 + \ell_3 \equiv 0 \pmod{\mathsf{p}} .$$

Hence the result. $\square$

Let G be a nontrivial irreducible graph with period p. Two states u and v in G are called *congruent*, denoted $u \equiv v$, if there is a path in G from u to v of length divisible by p. It can be readily verified that congruence is an equivalence relation that induces a partition on the states into equivalence classes.

Let C_0 be such an equivalence class, and for $r = 1, 2, \ldots, \mathsf{p}-1$, let C_r be the set of terminal states of edges in G whose initial states are in C_{r-1}, thus forming the sequence

$$C_0 \to C_1 \to \ldots \to C_{\mathsf{p}-1} .$$

The sets $C_0, C_1, \ldots, C_{\mathsf{p}-1}$ are necessarily all distinct, or else we would have a cycle in G whose length is less than p. The outgoing edges from $C_{\mathsf{p}-1}$ end path of length p that originate in C_0 and, so, their terminal states belong to C_0. It follows that the sets C_r form a partition of the set of states of G. In fact, each C_r is an equivalence of the congruence relation. Indeed, consider two states $u, v \in C_r$. There are paths in G,

$$u \to u_{r+1} \to u_{r+2} \to \ldots \to u_{\mathsf{p}-1} \to u_0 ,$$

and

$$v_0 \to v_1 \to v_2 \to \ldots \to v_{r-1} \to v ,$$

of lengths $\mathsf{p}-r$ and r, respectively, where $u_r, v_r \in C_r$; and since u_0 and v_0 are congruent, then so are u and v.

Example 3.5 The graph in Figure 3.2 has period 2, and the equivalence classes of the congruence relation are given by $C_0 = \{0, 2\}$ and $C_1 = \{1\}$. □

The definitions of period and primitivity extend to irreducible matrices through their support graphs as follows. Let A be a nontrivial irreducible matrix. The *period* $\mathsf{p} = \mathsf{p}(A)$ of A is the period of the support graph of A. A nontrivial irreducible matrix A is called *primitive* if the support graph of A is primitive.

Theorem 3.10 *Let A be a nontrivial irreducible matrix with period p and let*

$$C_0, C_1, \ldots, C_{\mathsf{p}-1}$$

be the equivalence classes of the congruence relation defined on the states of the support graph of A, where edges that start in C_r terminate in C_{r+1} (C_0 if $r = \mathsf{p}-1$).

(a) The nonzero entries of A all belong to p submatrices $B_0, B_1, \ldots, B_{\mathsf{p}-1}$ of A, where each B_r has order $|C_r| \times |C_{r+1}|$ ($|C_{\mathsf{p}-1}| \times |C_0|$ if $r = \mathsf{p}-1$).

(b) A^{p} decomposes into p irreducible components $A_0, A_1, \ldots, A_{\mathsf{p}-1}$, where

$$A_r = B_r B_{r+1} \cdots B_{\mathsf{p}-1} B_0 \cdots B_{r-1} .$$

Furthermore, the entries of A^{p} that do not belong to any of the irreducible components are all zero (i.e., the irreducible components of the support graph of A^{p} are isolated).

(c) Each irreducible component A_r of A^{p} is primitive.

(d) The irreducible components of A^{p} all have the same set of <u>nonzero</u> eigenvalues, with the same multiplicity.

We present below a (partial) proof of the theorem. The statement of the theorem can be seen more clearly if we apply the same permutation on the rows and columns of A so that for $r = 1, 2, \ldots, \mathsf{p}-1$, the states of C_r follow those of C_{r-1}. In such a case, A and A^{p} take the form

$$\begin{pmatrix} & \boxed{B_0} & & & \\ & & \boxed{B_1} & & \\ & & & \ddots & \\ & & & & \boxed{B_{\mathsf{p}-2}} \\ \boxed{B_{\mathsf{p}-1}} & & & & \end{pmatrix} \quad\text{and}\quad \begin{pmatrix} \boxed{A_0} & & & & \\ & \boxed{A_1} & & & \\ & & \ddots & & \\ & & & \boxed{A_{\mathsf{p}-2}} & \\ & & & & \boxed{A_{\mathsf{p}-1}} \end{pmatrix},$$

respectively. Note that the classes C_r need not necessarily be of the same size and the irreducible components A_r thus do not necessarily have the same order: different orders indicate different multiplicity of the zero eigenvalue.

Example 3.6 Continuing Example 3.5, consider again the graph G in Figure 3.2, which has period 2 and the equivalence classes of the congruence relation are $C_0 = \{0, 2\}$ and $C_1 = \{1\}$. As mentioned in Example 3.4, the adjacency matrix of G is given by

$$A_G = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

and after permuting the rows and columns of A_G so that the element(s) of C_1 follow those of C_0, we obtain the matrix

$$A = \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \hline 1 & 1 & 0 \end{array}\right) = \left(\begin{array}{c|c} 0 & B_0 \\ \hline B_1 & 0 \end{array}\right),$$

where

$$B_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

The second power of A is given by

$$A^2 = \left(\begin{array}{c|c} B_0 B_1 & 0 \\ \hline 0 & B_1 B_0 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \hline 0 & 0 & 2 \end{array}\right).$$

The irreducible components G_0 and G_1 of G^2 are shown in Figure 2.16, and their adjacency matrices are

$$A_0 = B_0 B_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad A_1 = B_1 B_0 = \begin{pmatrix} 2 \end{pmatrix} ,$$

respectively. The eigenvalues of A_0 are 2 and 0, out of which only 2 is an eigenvalue of A_1.□

Proof of Theorem 3.10. (a) follows from the definition of C_r, and the expression for A_r in (b) follows from the rules of matrix multiplication. It is left as an exercise (see Problem 3.10) to show that each A_r is irreducible and primitive and that the nonzero entries in A^{p} all belong to the submatrices A_r.

As for (d), let μ be a nonzero eigenvalue of A_r; that is, there exists a nonzero vector $\mathbf{x}$ such that

$$A_r \mathbf{x} = \mu \mathbf{x} .$$

Multiplying both sides by B_{r-1} we obtain

$$B_{r-1} A_r \mathbf{x} = \mu B_{r-1} \mathbf{x} .$$

Now, $B_{r-1} A_r = A_{r-1} B_{r-1}$; so,

$$A_{r-1}(B_{r-1}\mathbf{x}) = \mu (B_{r-1}\mathbf{x}) .$$

Furthermore, the vector $B_{r-1}\mathbf{x}$ is nonzero, or else we would have $A_r \mathbf{x} = B_r B_{r+1} \ldots B_{r-1}\mathbf{x} = \mathbf{0}$, contrary to our assumption that $\mu \neq 0$. Hence, it follows that μ is an eigenvalue of A_{r-1}. By perturbation it can be shown that μ has the same algebraic multiplicity as an eigenvalue of A_r and A_{r-1}. □

3.3.3 Perron-Frobenius Theorem

Theorem 3.11 (Perron-Frobenius Theorem for irreducible matrices) [Gant60, Ch. XIII], [Minc88, Ch. 1], [Sen80, Ch. 1], [Var62, Ch. 2]) *Let A be a nontrivial irreducible matrix. Then there exists an eigenvalue λ of A such that the following holds.*

(a) λ is real and $\lambda > 0$.

(b) There are right and left eigenvectors associated with λ that are strictly positive; that is, each of their components is strictly positive.

(c) $\lambda \geq |\mu|$ for any other eigenvalue μ of A.

(d) The geometric multiplicity of λ is 1; that is, the right and left eigenvectors associated with λ are unique up to scaling.

Proof. *Parts (a) and (b).* Let A be of order $m \times m$ and define the set

$$\mathcal{B} = \{\mathbf{y} \in \mathbb{R}^m \,:\, \mathbf{y} \geq \mathbf{0}\} \ .$$

For $\mathbf{y} = (y_u)_u \in \mathcal{B}$, let $\rho(\mathbf{y})$ be defined by

$$\rho(\mathbf{y}) = \min_{u \,:\, y_u > 0} \frac{(A\mathbf{y})_u}{y_u} \ .$$

Denoting by $u_{\max}$ the index u for which y_u is maximal, we have

$$0 \leq \rho(\mathbf{y}) \leq \frac{(A\mathbf{y})_{u_{\max}}}{y_{u_{\max}}} \leq \sum_v A_{u_{\max},v} \leq \max_u \sum_v A_{u,v} \ .$$

Therefore, the values $\rho(\mathbf{y})$ are uniformly bounded for every $\mathbf{y} \in \mathcal{B}$. Define

$$\lambda = \sup_{\mathbf{y} \in \mathcal{B}^*} \rho(\mathbf{y}) = \sup_{\mathbf{y} \in \mathcal{B}} \rho(\mathbf{y}) \ ,$$

where $\mathcal{B}^* = \{(y_u)_u \in \mathcal{B} \,:\, \sum_u y_u = 1\}$. Since the function $\mathbf{y} \mapsto \rho(\mathbf{y})$ is continuous over the compact set $\mathcal{B}^*$, there is some $\mathbf{x} \in \mathcal{B}^*$ for which $\rho(\mathbf{x}) = \lambda$. Observing that $A\mathbf{y} \geq \rho(\mathbf{y}) \cdot \mathbf{y}$ for every $\mathbf{y} \in \mathcal{B}$, it follows that

$$A\mathbf{x} \geq \lambda\mathbf{x} \ . \tag{3.2}$$

Next we show that the latter inequality holds with equality.

Suppose to the contrary that $A\mathbf{x} - \lambda\mathbf{x}$ is nonzero (and nonnegative). Define $B = (A + I)^{m-1}$, where I is the identity matrix; the matrix B is strictly positive (see Problem 3.14) and, therefore,

$$B(A\mathbf{x} - \lambda\mathbf{x}) > \mathbf{0} \ .$$

Letting $\mathbf{z} = (z_u)_u$ denote the vector $B\mathbf{x}$ and noting that B commutes with A, we have

$$A\mathbf{z} > \lambda\mathbf{z} \ ,$$

and, so,

$$\rho(\mathbf{z}) = \min_u \frac{(A\mathbf{z})_u}{z_u} > \lambda = \sup_{\mathbf{y} \in \mathcal{B}} \rho(\mathbf{y}) \ ,$$

thereby reaching a contradiction. We thus conclude that $A\mathbf{x} = \lambda\mathbf{x}$, i.e., λ is an eigenvalue of A with an associated nonnegative right eigenvector $\mathbf{x} = (x_u)_u$.

Next we show that both λ and $\mathbf{x}$ are strictly positive. Let the index v be such that $x_v > 0$, and for any index $u \neq v$, let $\ell_{u,v}$ be a positive integer for which $(A^{\ell_{u,v}})_{u,v} > 0$. Then, from $A^{\ell_{u,v}}\mathbf{x} = \lambda^{\ell_{u,v}}\mathbf{x}$ we obtain

$$\lambda^{\ell_{u,v}} x_u = (\lambda^{\ell_{u,v}}\mathbf{x})_u = (A^{\ell_{u,v}}\mathbf{x})_u \geq (A^{\ell_{u,v}})_{u,v} x_v > 0 \ .$$

Hence, $\lambda > 0$ and $\mathbf{x} > \mathbf{0}$. This completes the proof of part (a) and the first half of part (b): we still need to show that there is a strictly positive *left* eigenvector associated with λ. However, the existence of such a vector will follow from having, by (c), the same value of λ for the transpose of A.

Part (c). Let μ be a complex eigenvalue of A with an associated complex right eigenvector $\mathbf{y} = (y_u)_u$ and define the vector $\boldsymbol{\xi} = (\xi_u)_u$ by $\xi_u = |y_u|$. Taking the absolute value of both sides of

$$\sum_v (A)_{u,v} y_v = \mu \, y_u \ ,$$

we obtain, by the triangle inequality,

$$(A\boldsymbol{\xi})_u = \sum_v (A)_{u,v} |y_v| \geq \Big| \sum_v (A)_{u,v} y_v \Big| = |\mu| \xi_u \ ,$$

i.e.,

$$A\boldsymbol{\xi} \geq |\mu| \boldsymbol{\xi} \ .$$

Therefore,

$$|\mu| \leq \rho(\boldsymbol{\xi}) \leq \lambda \ . \tag{3.3}$$

Part (d). Let $\mathbf{x} = (x_u)_u$ be a strictly positive right eigenvector associated with the eigenvalue λ. Since λ is real, the linear space of the eigenvectors associated with λ is spanned by *real* eigenvectors. Let $\mathbf{y} = (y_u)_u$ be a real right eigenvector associated with λ and suppose to the contrary that $\mathbf{y}$ is linearly independent of $\mathbf{x}$. Then, for $\alpha = \max_u \{y_u / x_u\}$, the vector $\mathbf{z} = (z_u)_u = \alpha \mathbf{x} - \mathbf{y}$ is a nonnegative (nonzero) right eigenvector associated with λ and $z_u = 0$ for some index u. From $A\mathbf{z} = \lambda \mathbf{z}$ we obtain that $z_v = 0$ for every index v such that $(A)_{u,v} > 0$. Iterating inductively with each such v, we reach by the irreducibility of A the contradiction $\mathbf{z} = \mathbf{0}$. The respective proof for left eigenvectors is similar. $\square$

Hereafter, we denote the transpose of a vector $\mathbf{y}$ by $\mathbf{y}^\top$.

Proposition 3.12 *Let A and B be nonnegative real square submatrices of the same order such that $A \geq B$ and A is irreducible. Then, $\lambda(A) \geq \lambda(B)$, with equality if and only if $A = B$.*

Proof. Let $\mathbf{z} = (z_u)_u$ be a right eigenvector of B associated with an eigenvalue μ such that $|\mu| = \lambda_B = \lambda(B)$ and let $\mathbf{x} = (x_u)_u$ be defined by $x_u = |z_u|$; from $B\mathbf{z} = \mu \mathbf{z}$ and the triangle inequality we have

$$A\mathbf{x} \geq B\mathbf{x} \geq \lambda_B \mathbf{x} \ , \tag{3.4}$$

where the first inequality follows from $A \geq B$. Let $\mathbf{y}^\top$ be a strictly positive left eigenvector of A associated with $\lambda_A = \lambda(A)$. Multiplying by $\mathbf{y}^\top$ yields

$$\lambda_A \mathbf{y}^\top \mathbf{x} = \mathbf{y}^\top A \mathbf{x} \geq \mathbf{y}^\top B \mathbf{x} \geq \lambda_B \mathbf{y}^\top \ , \tag{3.5}$$

and dividing by the positive constant $\mathbf{y}^\top \mathbf{x}$, we obtain $\lambda_A \geq \lambda_B$.

Now, if $\lambda_A = \lambda_B$, then the inequalities in (3.5) must hold with equality. In fact, this is also true for the inequalities in (3.4), since $\mathbf{y}^\top$ is strictly positive; that is,

$$A\mathbf{x} = B\mathbf{x} = \lambda_B \mathbf{x} = \lambda_A \mathbf{x} \ .$$

It follows that $\mathbf{x}$ is a nonnegative right eigenvector of A associated with λ_A; as such, it must be strictly positive. Combining this with $B\mathbf{x} = A\mathbf{x}$ and $A \geq B$ yields $A = B$. $\square$

Proposition 3.13 *Let A be an irreducible matrix. Then the algebraic multiplicity of the eigenvalue $\lambda = \lambda(A)$ is 1; that is, λ is a simple root of the characteristic polynomial of A.*

Proof. The result is obvious for 1×1 matrices, so we exclude this case hereafter in the proof.

It is known that for every square matrix M,

$$M \cdot \mathrm{Adj}(M) = \det(M) \cdot I \ ,$$

where $\mathrm{Adj}(M)$ is the adjoint of M and I is the identity matrix. In particular,

$$(zI - A) \cdot \mathrm{Adj}(zI - A) = \chi_A(z) \cdot I \ ,$$

where $\chi_A(z) = \det(zI - A)$ is the characteristic polynomial of A. Differentiating with respect to z we obtain

$$\mathrm{Adj}(zI - A) + (zI - A) \cdot \frac{d}{dz}\Big(\mathrm{Adj}(zI - A)\Big) = \chi'_A(z) \cdot I \ .$$

We now substitute $z = \lambda$ and multiply each term by a strictly positive left eigenvector $\mathbf{y}^\top$ associated with λ; since $\mathbf{y}^\top(\lambda I - A) = \mathbf{0}^\top$, we end up with

$$\mathbf{y}^\top \mathrm{Adj}(\lambda I - A) = \chi'_A(\lambda)\, \mathbf{y}^\top \ .$$

Now, λ is a simple root of $\chi_A(z)$ if and only if $\chi'_A(\lambda) \neq 0$. Hence, to complete the proof, it suffices to show that the matrix $\mathrm{Adj}(\lambda I - A)$ is not all-zero. We do this next.

Let the matrix B be obtained from A by replacing the first row with the all-zero row. Denoting by $\chi_B(z)$ the characteristic polynomial of B, it is easy to see that the upper-left entry in $\mathrm{Adj}(\lambda I - A)$ is given by

$$\lambda^{-1} \chi_B(\lambda) \ .$$

However, from Proposition 3.12 it follows that λ is not an eigenvalue of B and, thus, cannot be a root of $\chi_B(z)$. $\square$

Proposition 3.14 *Let A be an irreducible matrix. Then,*

$$\min_u \sum_v (A)_{u,v} \le \lambda(A) \le \max_u \sum_v (A)_{u,v} \ ,$$

where equality in one side implies equality in the other.

Proof. Let $\mathbf{y}^\top = (y_v)_v$ be a strictly positive left eigenvector associated with $\lambda = \lambda(A)$. Then $\sum_u y_u (A)_{u,v} = \lambda y_v$ for every index v. Summing over v, we obtain,

$$\sum_u y_u \sum_v (A)_{u,v} = \lambda \sum_v y_v \ ,$$

or

$$\lambda = \frac{\sum_u y_u \sum_v (A)_{u,v}}{\sum_v y_v} \ .$$

That is, λ is a weighted average (over v) of the values $\sum_v (A)_{u,v}$. $\square$

Theorem 3.15 (Perron-Frobenius Theorem for nonnegative matrices.) *Let A be a nonnegative real square matrix. Then, the following holds.*

(a) The set of eigenvalues of A is the union (with multiplicity) of the sets of eigenvalues of the irreducible components of A.

(b) $\lambda(A)$ is an eigenvalue of A and there are nonnegative right and left eigenvectors associated with $\lambda(A)$.

Proof. Part (a) follows from the block-triangular form of Figure 3.3 (see the proof of Theorem 3.7). Part (b) is left as an exercise (see Problem 3.21). $\square$

Since $\lambda(A)$ is actually an eigenvalue of A, we will refer to $\lambda(A)$ as the *largest eigenvalue of A* or the *Perron eigenvalue of A*.

When all the irreducible components of a nonnegative real $m \times m$ matrix A are trivial, then all the eigenvalues of A are zero. In this case, the characteristic polynomial of A is given by $\chi_A(z) = z^m$. Such a matrix is called *nilpotent*. Notice that the support graph of a nilpotent matrix A does not contain cycles and, so, there is no path in that graph of length m. Therefore, $A^m = 0$, consistently with Caley-Hamilton Theorem that states that the all-zero matrix is obtained when a square matrix is substituted in its characteristic polynomial.

3.3.4 Stronger properties in the primitive case

The following proposition says that in the primitive case, the inequality in Theorem 3.11(c) is strict.

Proposition 3.16 *Let A be a primitive matrix with $\lambda(A) = \lambda$. Then $|\mu| < \lambda$ for every eigenvalue $\mu \neq \lambda$ of A.*

Proof. We use the notations $\mathbf{y} = (y_u)_u$, $\boldsymbol{\xi} = (|y_u|)_u$, and $\rho(\cdot)$ as in the proof of Theorem 3.11(c). If $|\mu| = \lambda$ then it follows from (3.3) that $\rho(\boldsymbol{\xi}) = \lambda$, i.e.,

$$A\boldsymbol{\xi} \geq \lambda\boldsymbol{\xi} \ .$$

Re-iterating the arguments in the proof of parts (a) and (b) of Theorem 3.11 (see (3.2)), we conclude that $\boldsymbol{\xi}$ is a right eigenvector associated with the eigenvalue λ. Therefore, for every positive integer ℓ and every index u,

$$\Big|\sum_v (A^\ell)_{u,v} y_v\Big| = |\mu|^\ell |y_u| = \sum_v (A^\ell)_{u,v} |y_v| \ ,$$

i.e., the triangle inequality holds with equality. In such a case we have for every v,

$$(A^\ell)_{u,v} y_v = (A^\ell)_{u,v} |y_v| \cdot \beta \ ,$$

where $\beta = \beta(u, \ell)$ is such that $|\beta| = 1$. Taking ℓ so that $A^\ell > 0$, we obtain that $\mathbf{y}$ is a scalar multiple of $\boldsymbol{\xi}$ and $\mu = \lambda$. $\square$

Theorem 3.17 *Let A be a primitive matrix and $\mathbf{x}$ and $\mathbf{y}^\top$ be strictly positive right and left eigenvectors of A associated with the eigenvalue $\lambda = \lambda(A)$, normalized so that $\mathbf{y}^\top \mathbf{x} = 1$. Then,*

$$\lim_{\ell \to \infty} (\lambda^{-1} A)^\ell = \mathbf{x}\,\mathbf{y}^\top \ .$$

Proof. The 1×1 case is immediate, so we exclude it from now on. Let μ be the largest absolute value of any eigenvalue of A other than λ; by Proposition 3.16 we have $\mu < \lambda$. Also, let h be the algebraic multiplicity of any eigenvalue of A whose absolute value equals μ. We show that

$$A^\ell = \lambda^\ell \mathbf{x}\mathbf{y}^\top + E^{(\ell)} \ ,$$

where $E^{(\ell)}$ is a matrix of the same order of A whose entries satisfy

$$|E^{(\ell)}_{u,v}| = O(\ell^{h-1} \mu^\ell) \tag{3.6}$$

for every u and v.

Write $A = P \Lambda P^{-1}$, where Λ is a matrix in *Jordan canonical form*; that is, Λ is a block-diagonal matrix where each block, Λ_i, is a square matrix that corresponds to an eigenvalue

λ_i and takes the *elementary Jordan form*: the entries on the main diagonal equal λ_i, and the other nonzero entries in the matrix are 1's below the main diagonal, as follows:

$$\Lambda_i = \begin{pmatrix} \lambda_i & & & & \\ 1 & \lambda_i & & & \\ & 1 & \lambda_i & & \\ & & \ddots & \ddots & \\ & & & 1 & \lambda_i \end{pmatrix} .$$

Each eigenvalue of A appears in the main diagonal of Λ a number of times which equals its algebraic multiplicity. We assume that the upper-left block Λ_1 corresponds to the largest eigenvalue λ; that is, $\Lambda_1 = \begin{pmatrix} \lambda \end{pmatrix}$ (by Proposition 3.13, this block has order 1×1). The first column of P and the first row of P^{-1} are, respectively, a right eigenvector $\mathbf{x}$ and a left eigenvector $\mathbf{y}^\top$ associated with λ, and $P \cdot P^{-1} = I$ implies that $\mathbf{x}$ and $\mathbf{y}^\top$ are normalized so that $\mathbf{y}^\top \mathbf{x} = 1$.

Now, $A^\ell = P\Lambda^\ell P^{-1}$, where Λ^ℓ is a block-diagonal matrix with blocks Λ_i^ℓ. It is easy to see that a each block Λ_i^ℓ is a lower-triangular matrix of the form

$$\Lambda_i^\ell = \begin{pmatrix} a_\ell & & & & \\ a_{\ell-1} & a_\ell & & & \\ a_{\ell-2} & a_{\ell-1} & a_\ell & & \\ \vdots & \ddots & \ddots & \ddots & \\ a_{\ell-s+1} & \ldots & a_{\ell-2} & a_{\ell-1} & a_\ell \end{pmatrix} ,$$

where s in the order of Λ_i and $a_j = \binom{\ell}{j}\lambda_i^j$. Hence, the absolute value of each entry in Λ_i^ℓ is bounded from above by $\ell^{s-1}|\lambda_i|^\ell$. Noting that $s \le h$, it follows that the upper-left entry of Λ^ℓ equals λ^ℓ, whereas the absolute values of the other entries of Λ^ℓ are bounded from above by $\ell^{h-1}\mu^\ell$ for sufficiently large ℓ. Therefore,

$$A^\ell = P\Lambda^\ell P^{-1} = \lambda^\ell \mathbf{x}\mathbf{y}^\top + E^{(\ell)} ,$$

where $E^{(\ell)}$ satisfies (3.6). □

When A is not primitive, there are eigenvalues of A other than λ for which Theorem 3.11(c) holds with equality. Those eigenvalues are identified in the next theorem, which is quoted here without proof.

Theorem 3.18 *Let A be a nontrivial irreducible matrix with period p and let $\lambda = \lambda(A)$. Then there are exactly p eigenvalues μ of A for which $|\mu| = \lambda$: those eigenvalues have the form $\lambda\omega^i$, where ω is a root of order p of unity, and each of those eigenvalues has algebraic multiplicity* 1.

3.4 Markov chains

In Section 3.1, we defined the capacity of a constrained system S combinatorially as the growth rate of the number of words in S. Then, in Section 3.2, we showed how the capacity of S was related to the largest eigenvalue of a lossless presentation of S. In Section 3.5 below, we present yet another characterization of capacity, now through probabilistic means.

The following concept plays a major role in our discussion.

Let $G = (V, E)$ be a graph. A *Markov chain on* G is a probability distribution $\mathcal{P}$ on the edges of G; namely, the mapping

$$e \mapsto \mathcal{P}(e)$$

takes nonnegative values and $\sum_{e \in E} \mathcal{P}(e) = 1$.

For a state $u \in V$, let E_u denote the set of outgoing edges from u in G, i.e.,

$$E_u = \{e \in E \,:\, \sigma(e) = u\} \ ,$$

where $\sigma(e) = \sigma_G(e)$ is the initial state of e in G. The *state probability vector* $\boldsymbol{\pi}^\top = (\pi_u)_{u \in V}$ of a Markov chain $\mathcal{P}$ on G is defined by

$$\pi_u = \sum_{e \in E_u} \mathcal{P}(e) \ .$$

The *conditional probability* of an edge $e \in E$ is defined by

$$q_e = \begin{cases} \mathcal{P}(e)/\pi_{\sigma(e)} & \text{if } \pi_{\sigma(e)} > 0 \\ 0 & \text{otherwise} \end{cases} \ .$$

A Markov chain $\mathcal{P}$ on G induces the following probability distribution on paths of G: given a path $\gamma = e_1 e_2 \ldots e_\ell$ in G, its probability is given by

$$\mathcal{P}(\gamma) = \pi_{\sigma(e_1)} q_{e_1} q_{e_2} \cdots q_{e_\ell} \ . \tag{3.7}$$

The *transition matrix* associated with $\mathcal{P}$ is a nonnegative real $|V| \times |V|$ matrix Q where for every $u, v \in V$,

$$(Q)_{u,v} = \sum_{e \in E_u \,:\, \tau(e) = v} q_e \ ;$$

that is, $(Q)_{u,v}$ is the sum of the conditional probabilities of all edges from u to v in G. Note that Q is *stochastic*: the sum of entries in each row is 1.

A Markov chain $\mathcal{P}$ on G is called *stationary* if for every $u \in V$,

$$\sum_{e \in E \,:\, \tau(e) = u} \mathcal{P}(e) = \pi_u \ ;$$

that is, the sum of the probabilities of the incoming edges to state u equals the respective sum of the outgoing edges from u. Equivalently,

$$\boldsymbol{\pi}^\top Q = \boldsymbol{\pi}^\top .$$

A stationary Markov chain $\mathcal{P}$ on G is called *irreducible* (or *ergodic*) if the associated transition matrix Q is irreducible. Similarly, $\mathcal{P}$ is called *primitive* (or *mixing*) if Q is a primitive matrix. Clearly, Q is irreducible (respectively, primitive) only if G is. Hereafter, when we say an irreducible (respectively, primitive) Markov chain, we mean an irreducible (respectively, primitive) *stationary* Markov chain.

Proposition 3.19 *Let Q be an irreducible stochastic $|V| \times |V|$ matrix. Then there is a unique positive vector $\boldsymbol{\pi}^\top = (\pi_u)_{u \in V}$ such that $\sum_u \pi_u = 1$ and*

$$\boldsymbol{\pi}^\top Q = \boldsymbol{\pi}^\top .$$

Proof. The matrix Q is irreducible and the sum of elements in each row is 1. By Proposition 3.14 we thus have $\lambda(Q) = 1$. The existence and uniqueness of $\boldsymbol{\pi}^\top$ now follow from parts (b) and (d) of Theorem 3.11. □

It follows from Proposition 3.19 that an irreducible Markov chain on $G = (V, E)$ is uniquely determined by its conditional edge probabilities $(q_e)_{e \in E}$. That is, these conditional probabilities determine the state probability vector. We refer to the state probability vector of an irreducible Markov chain as the *stationary probability vector*.

The *entropy* (or *entropy rate*) of a Markov chain $\mathcal{P}$ on $G = (V, E)$ is defined as the expected value—with respect to the probability measure $\mathcal{P}$ on the edges of G—of the random variable $\log(1/q_e)$; i.e.,

$$\mathsf{H}(\mathcal{P}) = \mathsf{E}_{\mathcal{P}} \left\{ \log(1/q_e) \right\} = -\sum_{u \in V} \pi_u \sum_{e \in E_u} q_e \log q_e .$$

Example 3.7 Let G be the Shannon cover of the $(0, 1)$-RLL constrained system as shown in Figure 3.1, and consider the following stochastic matrix (whose support graph is G):

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} .$$

One can verify that

$$\boldsymbol{\pi}^\top = \left(\tfrac{2}{3} \;\; \tfrac{1}{3} \right)$$

is a left eigenvector of Q associated with the Perron eigenvalue 1. The vector $\boldsymbol{\pi}^\top$ is the stationary probability vector of the (unique) stationary Markov chain on G whose transition matrix is Q. The entropy of this Markov chain is $-\sum_u \pi_u \sum_{e \in E_u} q_e \log q_e = 2/3$. □

Proposition 3.20 *Let Q be a primitive stochastic $|V| \times |V|$ matrix and let $\boldsymbol{\xi}^\top = (\xi_u)_{u \in V}$ be such that $\sum_{u \in V} \xi_u = 1$. Then,*

$$\lim_{\ell \to \infty} \boldsymbol{\xi}^\top Q^\ell = \boldsymbol{\pi}^\top ,$$

where $\boldsymbol{\pi}^\top$ is the vector as in Proposition 3.19.

Proof. Since Q is stochastic we have $Q\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-one column vector; that is, $\mathbf{1}$ is a right eigenvector associated with the Perron eigenvalue $\lambda(Q) = 1$. Hence, by Theorem 3.17 we have

$$\lim_{\ell \to \infty} \boldsymbol{\xi}^\top Q^\ell = \boldsymbol{\xi}^\top \mathbf{1}\, \boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top ,$$

as claimed. □

Suppose that $\mathcal{P}$ is a (not necessarily stationary) Markov chain on G with an associated primitive transition matrix Q and a state probability vector $\boldsymbol{\xi}^\top$. Also, let $\boldsymbol{\pi}^\top = (\pi_u)_u$ be a left positive eigenvector of Q associated with the Perron eigenvalue 1, normalized so that $\sum_u \pi_u = 1$. It follows from Proposition 3.20 that as the lengths of paths go to infinity, the probability of terminating in state u of G converges to π_u.

Theorem 3.21 (Law of large numbers for irreducible Markov chains) *Let $\mathcal{P}$ be an irreducible Markov chain on a labeled graph $G = (V, E, L)$ where $L : E \to \mathbb{R}$ (i.e., the labels are over the real field). For a positive integer ℓ, define the random variable Z_ℓ on paths $\gamma = e_1 e_2 \ldots e_\ell$ of length ℓ in G by*

$$Z_\ell = Z_\ell(\gamma) = \frac{1}{\ell} \sum_{i=1}^{\ell} L(e_i) .$$

Then, for every $\epsilon > 0$,

$$\lim_{\ell \to \infty} \mathsf{Prob}\Big\{ |Z_\ell - \overline{L}\,| \le \epsilon \Big\} = 1 ,$$

where $\overline{L} = \mathsf{E}_{\mathcal{P}}\{L(e)\} = \sum_{e \in E} \mathcal{P}(e) L(e)$.

The proof of Theorem 3.21 is left as a guided exercise (see Problems 3.35 and 3.36). Observe that since $\mathcal{P}$ is stationary, we have in fact $\overline{L} = \mathsf{E}_{\mathcal{P}}\{Z_\ell\}$.

Let $\mathcal{P}$ be a Markov chain on G. A path γ in G of length ℓ is called $(\mathcal{P}, \epsilon)$-*typical* if the probability $\mathcal{P}(\gamma)$, as defined by (3.7), satisfies

$$\left| \mathsf{H}(\mathcal{P}) + \frac{1}{\ell} \log \mathcal{P}(\gamma) \right| \le \epsilon ,$$

or, equivalently,

$$2^{-\ell(\mathsf{H}(\mathcal{P})+\epsilon)} \le \mathcal{P}(\gamma) \le 2^{-\ell(\mathsf{H}(\mathcal{P})-\epsilon)} .$$

The set of $(\mathcal{P}, \epsilon)$-typical paths of length ℓ in G will be denoted by $\mathcal{T}_\ell(\mathcal{P}, \epsilon)$ (the dependency of this set on G is implied by the dependency on $\mathcal{P}$).

Theorem 3.22 (Asymptotic Equipartition Property, in short AEP) *Let $\mathcal{P}$ be an irreducible Markov chain on a graph G. Then, for every $\epsilon > 0$,*

$$\lim_{\ell\to\infty} \sum_{\gamma\in\mathcal{T}_\ell(\mathcal{P},\epsilon)} \mathcal{P}(\gamma) = 1 \ .$$

Proof. We apply Theorem 3.21 to the graph G and the labeling $L(e) = \log(1/q_e)$, where q_e is the conditional probability of an edge e; we assume that $q_e > 0$, or else we delete the edge e from G. Here,

$$\overline{L} = \mathsf{E}_{\mathcal{P}}\{\log(1/q_e)\} = \mathsf{H}(\mathcal{P})$$

and, therefore,

$$\lim_{\ell\to\infty} \mathsf{Prob}\{\,|Z_\ell - \mathsf{H}(\mathcal{P})\,| \le \epsilon\,\} = 1 \ . \tag{3.8}$$

On the other hand, letting $(\pi_u)_u$ be the stationary probability vector of $\mathcal{P}$, we have for every path $\gamma = e_1e_2\ldots e_\ell$ in G,

$$-\frac{1}{\ell}\log\mathcal{P}(\gamma) = \frac{\sum_{i=1}^{\ell}\log(1/q_{e_i})}{\ell} + \frac{\log(1/\pi_{\sigma(e_1)})}{\ell} = Z_\ell(\gamma) + o(1) \ ,$$

where $o(1)$ stands for an expression that goes to zero as ℓ goes to infinity. Hence, by (3.8) we obtain

$$\lim_{\ell\to\infty} \sum_{\gamma\in\mathcal{T}_\ell(\mathcal{P},\epsilon)} \mathcal{P}(\gamma) = \lim_{\ell\to\infty} \mathsf{Prob}\left\{\,\left|\mathsf{H}(\mathcal{P}) + \frac{1}{\ell}\log\mathcal{P}(\gamma)\right| \le \epsilon\right\} = 1 \ ,$$

as claimed. □

The AEP thus states that for large ℓ, 'most' paths of length ℓ in G have probability roughly $2^{-\ell\mathsf{H}(\mathcal{P})}$; here, the quantifier 'most' does not refer to the actual count of the paths, but rather to their measure as induced by the probability distribution $\mathcal{P}$.

We remark that Theorem 3.22 holds also for the following stronger property of paths. Let $\mathcal{P}$ be a Markov chain on G. A path γ in G of length ℓ is called $(\mathcal{P},\epsilon)$-*strongly-typical* if for every edge e in G, the number of times that e is traversed in γ is bounded from below by $\ell(\mathcal{P}(e){-}\epsilon)$ and from above by $\ell(\mathcal{P}(e){+}\epsilon)$. It can shown that if $\mathcal{P}$ is irreducible, then this property implies that γ is typical (see Problem 3.37). The re-statement of Theorem 3.22 for strongly-typical paths is left as an exercise (Problem 3.38).

3.5 Probabilistic characterization of capacity

Theorem 3.23 *Let S be a constrained system which is presented by an irreducible lossless graph G. Then,*

$$\sup_{\mathcal{P}} \mathsf{H}(\mathcal{P}) = \log\lambda(A_G) = \mathsf{cap}(S) \ ,$$

where the supremum is taken over all stationary Markov chains on G.

Proof. By continuity, every stationary Markov chain $\mathcal{P}$ on G can be expressed as a limit of irreducible Markov chains $\mathcal{P}_1, \mathcal{P}_2, \cdots$ on G, and $\lim_{i\to\infty} \mathsf{H}(\mathcal{P}_i) = \mathsf{H}(\mathcal{P})$. Hence, it suffices to prove the theorem for irreducible Markov chains on G.

By Theorem 3.22, for every $\epsilon, \delta > 0$ there is a positive integer N such that for every $\ell \geq N$,

$$\sum_{\gamma \in \mathcal{T}_\ell(\mathcal{P},\epsilon)} \mathcal{P}(\gamma) \geq 1 - \delta \ .$$

On the other hand, for every $\gamma \in \mathcal{T}_\ell(\mathcal{P},\epsilon)$ we have

$$\mathcal{P}(\gamma) \leq 2^{-\ell(\mathsf{H}(\mathcal{P})-\epsilon)} \ .$$

Summing over $\gamma \in \mathcal{T}_\ell(\mathcal{P},\epsilon)$ yields

$$1 - \delta \leq \sum_{\gamma \in \mathcal{T}_\ell(\mathcal{P},\epsilon)} \mathcal{P}(\gamma) \leq |\mathcal{T}_\ell(\mathcal{P},\epsilon)| \cdot 2^{-\ell(\mathsf{H}(\mathcal{P})-\epsilon)} \ ,$$

or

$$|\mathcal{T}_\ell(\mathcal{P},\epsilon)| \geq (1-\delta) \cdot 2^{\ell(\mathsf{H}(\mathcal{P})-\epsilon)} \ .$$

Assuming that $\delta < 1$, by Lemma 3.5 we obtain

$$\mathsf{H}(\mathcal{P}) - \epsilon \leq \log \lambda(A_G)$$

for every $\epsilon > 0$. Hence, $\mathsf{H}(\mathcal{P}) \leq \log \lambda(A_G)$. To complete the proof, it suffices to exhibit a stationary Markov chain $\mathcal{P}$ on $G = (V,E)$ for which $\mathsf{H}(\mathcal{P}) = \log \lambda(A_G)$.

Let $\mathbf{x} = (x_u)_{u\in V}$ and $\mathbf{y}^\top = (y_u)_{u \in V}$ be positive right and left eigenvectors of A_G associated with the eigenvalue $\lambda = \lambda(A_G)$ and normalized so that $\mathbf{y}^\top \mathbf{x} = 1$. Define a stationary Markov chain $\mathcal{P}$ through the conditional probabilities

$$q_e = \frac{x_{\tau(e)}}{\lambda x_{\sigma(e)}} \ .$$

The entries of the transition matrix Q are given by

$$(Q)_{u,v} = \sum_{e \in E_u \,:\, \tau(e)=v} q_e = \frac{(A_G)_{u,v} x_v}{\lambda x_u} \ , \tag{3.9}$$

and it is easy to verify that Q is, indeed, a stochastic matrix on G. A simple computation shows that the vector $\boldsymbol{\pi}^\top = (\, x_u y_u \,)_{u \in V}$ satisfies $\boldsymbol{\pi}^\top Q = \boldsymbol{\pi}^\top$ and, so, it is the stationary probability vector of $\mathcal{P}$.

Now,

$$\mathsf{H}(\mathcal{P}) \;=\; -\sum_{u\in V} \pi_u \sum_{e \in E_u} q_e \log q_e$$

$$
\begin{aligned}
&= \sum_{u\in V} \pi_u \sum_{e\in E_u} q_e\Big((\log\lambda) + (\log x_u) - (\log x_{\tau(e)})\Big) \\
&= \underbrace{(\log\lambda) \sum_{u\in V} \pi_u \underbrace{\sum_{e\in E_u} q_e}_{1}}_{1} + \Big(\sum_{u\in V} (\pi_u \log x_u) \underbrace{\sum_{e\in E_u} q_e}_{1}\Big) - \Big(\sum_{u\in V} \pi_u \sum_{v\in V} (\log x_v) \underbrace{\sum_{e\in E_u\,:\,\tau(e)=v} q_e}_{(Q)_{u,v}}\Big) \\
&= (\log\lambda) + \sum_{u\in V} (\pi_u \log x_u) - \sum_{v\in V} (\log x_v) \underbrace{\sum_{u\in V} \pi_u (Q)_{u,v}}_{\pi_v} \\
&= \log\lambda\ .
\end{aligned}
$$

See also [Imm91, p. 48]. □

A stationary Markov chain $\mathcal{P}$ on G for which $\mathsf{H}(\mathcal{P}) = \log\lambda(A_G)$ is called a *maxentropic Markov chain on* G. We have shown in the proof of Theorem 3.23 that a maxentropic Markov chain exists for every irreducible graph. In fact, it can be shown that such a stationary Markov chain is unique [Par64] (see also [PT82]). It follows from Theorem 3.23 that if $\mathcal{P}$ is a maxentropic Markov chain on an irreducible graph G, then for every $\epsilon > 0$, the growth rate of the $(\mathcal{P}, \epsilon)$-typical paths in G is essentially the same as the number of paths in G. We can thus say that the $(\mathcal{P}, \epsilon)$-typical paths form a 'substantial subset' within the set of all paths in G. The same can be said about the $(\mathcal{P}, \epsilon)$-strongly-typical paths in G and, as such, a maxentropic Markov chain defines the frequency with which a symbol appears in a 'substantial subset' of words of $S(G)$. For the analysis of such frequency in (d, k)-RLL constrained systems, see [How89].

Example 3.8 Let G be the Shannon cover of the $(0, 1)$-RLL constrained system, as shown in Figure 3.1. The adjacency matrix of G is

$$
A_G = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} ,
$$

with Perron eigenvalue $\lambda = (1+\sqrt{5})/2 \approx 1.618$ and an associated right eigenvector

$$
\mathbf{x} = \begin{pmatrix} \lambda \\ 1 \end{pmatrix} .
$$

By (3.9), the transition matrix Q of the maxentropic Markov chain on G is given by

$$
Q = \begin{pmatrix} 1/\lambda & 1/\lambda^2 \\ 1 & 0 \end{pmatrix} \approx \begin{pmatrix} .618 & .382 \\ 1 & 0 \end{pmatrix}
$$

and the respective stationary probability vector is

$$
\boldsymbol{\pi}^\top = \left(\tfrac{\lambda+1}{\lambda+2} \;\; \tfrac{1}{\lambda+2} \right) \approx (\,.724 \;\; .276\,) .
$$

This means that in a 'substantial subset' of words of the $(0,1)$-RLL constrained system, approximately 27.6% of the symbols are 0.

The entropy of the maxentropic Markov chain is $\log \lambda \approx .6942$. Note that the stationary Markov chain in Example 3.7 has smaller entropy. □

3.6 Approaching capacity by finite-type constraints

The next two propositions exhibit an important feature of finite-type constrained systems.

Proposition 3.24 *Let S be a constrained system. Then, there is a sequence of finite-type constrained systems $\{S_m\}_{m=1}^{\infty}$ such that $S \subseteq S_m$ for every m and $\lim_{m\to\infty} \mathsf{cap}(S_m) = \mathsf{cap}(S)$.*

Proof. Given a positive integer m, we let S_m be the constrained system which is presented by the following deterministic graph G_m: For each word $\mathbf{w}$ of length m in S, we associate a state $u_{\mathbf{w}}$ in G_m. Given two words of length m in S, $\mathbf{w} = w_1 w_2 \ldots w_m$ and $\mathbf{z} = z_1 z_2 \ldots z_m$, we draw an edge $u_{\mathbf{w}} \xrightarrow{b} u_{\mathbf{z}}$ in G_m if and only if $b = z_m$ and $z_j = w_{j+1}$ for $j = 1, 2, \ldots, m-1$.

It is easy to verify that all paths in G_m that generate a word $\mathbf{w}$ terminate in $u_{\mathbf{w}}$. Hence, G_m has memory $\le m$. To show that $S \subseteq S_m$, let $\mathbf{z} = z_1 z_2 \ldots z_\ell$ be a word of length $\ell \ge m$ in S. Then, by construction, the word $\mathbf{z}$ is generated in G_m by a path

$$u_{w_1 w_2 \ldots w_m} \xrightarrow{z_1} u_{w_2 w_3 \ldots w_m z_1} \xrightarrow{z_2} u_{w_3 w_4 \ldots z_m z_1 z_2} \xrightarrow{z_3} \cdots \xrightarrow{z_\ell} u_{z_{\ell-m+1} z_{\ell-m+2} \ldots z_\ell} .$$

Hence, $\mathbf{z}$ is a word in S_m. Furthermore, we also have $S_{m+1} \subseteq S_m$ and, so, the values of $\mathsf{cap}(S_m)$ are nonincreasing with m. Therefore, the limit $\lim_{m\to\infty} \mathsf{cap}(S_m)$ exists and it is at least $\mathsf{cap}(S)$. It remains to show that it is actually equal to $\mathsf{cap}(S)$.

Let $N(m; S)$ be the number of words of length m in S. Every word of length tm in S_m can be written as a concatenation of t words of length m in S. Therefore, the number of words of length tm in S_m is at most $(N(m; S))^t$. Thus, $\mathsf{cap}(S_m) \le (\log N(m; S))/m$ and, so,

$$\mathsf{cap}(S) \le \lim_{m\to\infty} \mathsf{cap}(S_m) \le \lim_{m\to\infty} (\log N(m; S))/m = \mathsf{cap}(S) ,$$

as desired. □

The following dual result is proved in [Mar85].

Proposition 3.25 *Let S be a constrained system. Then, there is a sequence of finite-type constrained systems $\{S_m\}_{m=1}^{\infty}$ such that $S_m \subseteq S$ for every m and $\lim_{m\to\infty} \mathsf{cap}(S_m) = \mathsf{cap}(S)$.*

Sketch of proof (in the irreducible case): Let S_m be the constrained system which is presented by the following deterministic graph G_m: For every magic word $\mathbf{w}$ of length m in S (see Section 2.6.4), we associate a state $u_{\mathbf{w}}$ in G_m. Given two magic words of length m in S, $\mathbf{w} = w_1 w_2 \ldots w_m$ and $\mathbf{z} = z_1 z_2 \ldots z_m$, we draw an edge $u_{\mathbf{w}} \xrightarrow{b} u_{\mathbf{z}}$ in G_m if and only if the following three conditions hold:

(a) $z_j = w_{j+1}$ for $j = 1, 2, \ldots, m-1$;

(b) $b = z_m$;

(c) $\mathbf{w}b \in S$.

It is easy to verify that G_m has memory $\leq m$ and that $S_m \subseteq S$. The approach of $\mathsf{cap}(S_m)$ to $\mathsf{cap}(S)$ follows from the fact that most long enough words in S are magic words. The precise proof is given in [Mar85]. $\square$

Problems

Problem 3.1 Let S be a constrained system and ℓ a positive integer. Based on the definition of capacity, show that

$$\mathsf{cap}(S^\ell) = \ell \cdot \mathsf{cap}(S) \ .$$

The following is a skeleton of a proof for the inequality $\mathsf{cap}(S^\ell) \geq \ell \cdot \mathsf{cap}(S)$. Justify each step.

Denote by Σ the alphabet of S and let $\ell_1 < \ell_2 < \cdots < \ell_i < \cdots$ be such that

$$\lim_{i\to\infty} \frac{1}{\ell_i} \log N(\ell_i; S) = \mathsf{cap}(S)$$

(why do such ℓ_i's exist?). Define $m_i = \lfloor \ell_i/\ell \rfloor$. Then

$$\begin{aligned}
\mathsf{cap}(S^\ell) &\geq \limsup_{i\to\infty} \frac{1}{m_i} \log N(m_i \ell; S) \\
&\geq \limsup_{i\to\infty} \frac{1}{m_i} \log N(\ell_i; S) \\
&\geq \ell \cdot \lim_{i\to\infty} \frac{\ell_i}{m_i \ell} \lim_{i\to\infty} \frac{1}{\ell_i} \log N(\ell_i, S) \\
&= \ell \cdot \mathsf{cap}(S) \ .
\end{aligned}$$

Problem 3.2 Let $\varepsilon : \mathbb{N} \to \mathbb{R}^+$ be a function such that $\lim_{m\to\infty} \varepsilon(m) = 0$ and for a positive integer h define

$$\delta(\ell) = \frac{1}{\ell} \max_{(\ell_1,\ell_2,\ldots,\ell_h)} \sum_{i=1}^{h} \ell_i \varepsilon(\ell_i) \ ,$$

where $(\ell_1, \ell_2, \ldots, \ell_h)$ ranges over all nonnegative integer h-tuples such that $\ell_1 + \ell_2 + \ldots + \ell_h \leq \ell$. Complete the proof of Theorem 3.1 by showing that $\lim_{\ell\to\infty} \delta(\ell) = 0$.

Hint: Let β be a finite upper bound on the values of $\varepsilon(m)$. Justify each of the following steps:

$$\begin{aligned}
\delta(\ell) &= \max_{(\ell_1,\ell_2,\ldots,\ell_h)} \sum_{i=1}^{h} \frac{\ell_i}{\ell}\,\varepsilon(\ell_i) \\
&= \max_{(\ell_1,\ell_2,\ldots,\ell_h)} \left(\left(\sum_{i\,:\,\ell_i \le \sqrt{\ell}} \frac{\ell_i}{\ell}\,\varepsilon(\ell_i) \right) + \left(\sum_{i\,:\,\ell_i > \sqrt{\ell}} \frac{\ell_i}{\ell}\,\varepsilon(\ell_i) \right) \right) \\
&\le \frac{\beta \cdot h}{\sqrt{\ell}} + \max_{(\ell_1,\ell_2,\ldots,\ell_h)} \left(\sum_{i\,:\,\ell_i > \sqrt{\ell}} \varepsilon(\ell_i) \right) ,
\end{aligned}$$

and then deduce that $\lim_{\ell \to \infty} \delta(\ell) = 0$.

Problem 3.3 Let $S_{d,\infty}$ denote the (d,∞)-RLL constrained system.

1. Show that
$$\mathsf{cap}(S_{d,\infty}) \le \frac{\log(d+2)}{d+1} \ .$$
Hint: Show that when a word in $S_{d,\infty}$ is divided into nonoverlapping blocks of length $d+1$, then each such block may contain at most one 1.

2. Show that for every positive integer m,
$$\mathsf{cap}(S_{d,\infty}) \ge \frac{\log(m+1)}{d+m} \ .$$
Hint: Consider the concatenation of binary blocks of length $m+d$, each containing at most one 1 which is located in one of the first m positions of the block.

3. Show that
$$\lim_{d\to\infty} \mathsf{cap}(S_{d,\infty}) \cdot \frac{d}{\log d} = 1 \ .$$
Hint: Substitute $m = \varepsilon d$ in 2 and let $d \to \infty$ for every fixed small ε.

Problem 3.4 Compute the capacity of the $(d,\infty,2)$-RLL constraint for $d = 0, 1$; recall that this constraint consist of all binary words in which the runlengths of 0's between consecutive 1's are even when $d = 0$ and odd when $d = 1$.

Problem 3.5 Identify the values of d and k for which the Shannon cover in Figure 1.3 is periodic.

Problem 3.6 Let G be a nontrivial irreducible graph and let G' and G'' be the Moore form and Moore co-form of G, respectively. Show that the periods of G, G', and G'' are equal.

Problem 3.7 Let G_1 and G_2 be nontrivial irreducible graphs with periods p_1 and p_2, respectively. Show that the period of each *nontrivial* irreducible component of $G_1 * G_2$ is divisible by the least common multiplier (l.c.m.) of p_1 and p_2.

Problem 3.8 A path

$$u_0 \to u_1 \to u_2 \to \cdots \to u_\ell$$

in a graph is called a *simple cycle* if $u_0 = u_\ell$ and $u_i \neq u_j$ for $0 \leq i < j < \ell$. Let G be a nontrivial irreducible graph with period p. Show that p is the greatest common divisor of the lengths of all the simple cycles in G.

Problem 3.9 Let G be a nontrivial irreducible graph with period p and let v be a state in G. Show that p is the greatest common divisor of the lengths of all cycles in G that pass through v.

Problem 3.10 Let G be a nontrivial irreducible graph with period p.

1. Show that for every pair of states u and v in G there exist nonnegative integers $m_{u,v}$ and $r_{u,v}$, such that for every integer $m \geq m_{u,v}$ there is a path in G of length $m \cdot \mathsf{p} + r_{u,v}$ originating in state u and terminating in v.
2. Show that G^{p} has p irreducible components which are primitive and isolated from each other.
3. Let ℓ be a positive integer relatively prime to p (i.e., $\gcd(\mathsf{p}, \ell) = 1$). Show that G^ℓ is irreducible.

   Hint: Make use of 1 and the fact that there is a positive integer b such that $b \cdot \mathsf{p} \equiv 1 \pmod{\ell}$.
4. Generalize 2 and 3 as follows. Let ℓ be a positive integer and $d = \gcd(\mathsf{p}, \ell)$. Show that G^ℓ composes of d isolated irreducible components, each with period p/d.

Problem 3.11 Let G be a nontrivial irreducible (not necessarily lossless) graph with period p and let $G_0, G_1, \ldots, G_{\mathsf{p}-1}$ be the irreducible components of G^{p}. Show from the definition of capacity that

$$\mathsf{cap}(S(G_i)) = \mathsf{p} \cdot \mathsf{cap}(S(G))$$

for every irreducible component G_i.

Problem 3.12 Let A be a nonnegative real square matrix and let ℓ be a positive integer. Show that the irreducible components of A^ℓ all have the same set of nonzero eigenvalues, with the same multiplicity.

Problem 3.13 Let A be the matrix given by

$$A = \begin{pmatrix} 2 & 5 & 0 \\ 0 & 1 & 6 \\ 1 & 3 & 3 \end{pmatrix} .$$

1. Compute the eigenvalues of A and respective left and right *integer* eigenvectors.
2. Find a diagonal matrix Λ and an invertible matrix P such that $A = P\Lambda P^{-1}$.
3. Compute the limit
$$B = \lim_{\ell \to \infty} \frac{1}{7^\ell} \cdot A^\ell .$$
4. Find integer vectors that span the row space and column space, respectively, of B.

Problem 3.14 Let A be a nonnegative real square matrix of order $m \times m$. Show that A is irreducible if and only if $(A+I)^{m-1} > 0$.

Problem 3.15 Let G be a graph with an adjacency matrix

$$A_G = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} ,$$

1. What is the period p of A_G?
2. Write the matrix A_G^{p}.
3. Compute the absolute values of the eigenvalues of A_G.

Problem 3.16 Let G be the graph in Figure 2.24.

1. What is the period p of G?
2. Find the irreducible components G_i of G^{p}.
3. For every component G_i found in 2, compute the value of $\lambda(A_{G_i})$.
4. Compute $\lambda(A_G)$.
5. For every component G_i found in 2, compute a positive *integer* eigenvector associated with $\lambda(A_{G_i})$, such that the largest entry in the vector is the smallest possible.
6. Compute an eigenvector $\mathbf{x} = (x_{u_i})_{i=1}^8$ of A_G associated with $\lambda(A_G)$ such that $x_{u_1} = 1$. How many such vectors exist?
7. Is there an *integer* eigenvector of A_G associated with $\lambda(A_G)$?
8. Repeat 6 except that now $\mathbf{x}$ is an eigenvector of A_{G^2} associated with $\lambda(A_{G^2})$.
9. Compute a nonnegative integer eigenvector of A_{G^2} associated with $\lambda(A_{G^2})$ where the largest entry in that eigenvector is the smallest possible.

Problem 3.17 Let $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ be two labeled graphs with labeling $L_1 : E_1 \to \Sigma_1$ and $L_2 : E_2 \to \Sigma_2$. The *Kronecker product* of G_1 and G_2 is the labeled graph

$$G = G_1 \otimes G_2 = (V_1 \times V_2, E, L) \ ,$$

where the set of edges E and the labeling $L : E \to \Sigma_1 \times \Sigma_2$ are defined as follows:

$$(u_1, u_2) \stackrel{(a_1,a_2)}{\longrightarrow} (v_1, v_2)$$

is an edge in E if and only if

$$u_1 \stackrel{a_1}{\to} v_1 \quad \text{and} \quad u_2 \stackrel{a_2}{\to} v_2$$

are edges in G_1 and G_2, respectively.

Let S_1, S_2, and S be the constrained systems that are generated by G_1, G_2, and G, respectively.

Hereafter assume that G_1 and G_2 are nontrivial irreducible graphs and let p_1 and p_2 denote their periods, respectively.

1. Show that
$$\mathsf{cap}(S) = \mathsf{cap}(S_1) + \mathsf{cap}(S_2)$$
(do not assume that G_1 and G_2 are lossless).

2. Show that the anticipation of G is given by
$$\mathcal{A}(G) = \max\Big\{\mathcal{A}(G_1), \mathcal{A}(G_2)\Big\} \ .$$

3. Show that if G_1 and G_2 are deterministic and have finite memory, then the memory of G is given by
$$\mathcal{M}(G) = \max\{\mathcal{M}(G_1), \mathcal{M}(G_2)\} \ .$$

4. Show that G has $\gcd(\mathsf{p}_1, \mathsf{p}_2)$ irreducible components, which are isolated from each other.

5. Show that the period of each irreducible component of G is the least common multiplier (lcm) of p_1 and p_2.

6. Show that the adjacency matrix A_G is the Kronecker (or direct) product of A_{G_1} and A_{G_2}, namely, for every (u_1, u_2) and (v_1, v_2) in V,
$$(A_G)_{(u_1,u_2),(v_1,v_2)} = (A_{G_1})_{u_1,v_1} \cdot (A_{G_2})_{u_2,v_2} \ .$$

7. Show that
$$\lambda(A_G) = \lambda(A_{G_1}) \cdot \lambda(A_{G_2})$$
(this is a known property of Kronecker product of matrices, but it can be proved also by counting paths in G_1, G_2, and G).

Problem 3.18 Let $G = (V, E, L)$ be a primitive graph and let $\lambda = \lambda(A_G)$. Denote by μ the largest absolute value of any eigenvalue of A_G other than λ (let $\mu = 0$ if $|V| = 1$). Show that the number, $\Phi_G(\ell)$, of cycles of length ℓ in G satisfies

$$|\Phi_G(\ell) - \lambda^\ell| \le (|V| - 1)\mu^\ell .$$

Hint: Recall that the trace of a square matrix B, which is the sum of the entries along the main diagonal of B, is preserved in the Jordan form of B (as well as in any other matrix that is similar to B).

Problem 3.19 For $0 \le d \le k < \infty$, let $G(d, k)$ be the Shannon cover in Figure 1.3 of the (d, k)-RLL constrained system and let $\chi_{d,k}(z) = \det(zI - A_{G(d,k)})$ be the characteristic polynomial of $A_{G(d,k)}$.

1. Show that
$$\chi_{d,k}(z) = z^{k+1} - \sum_{j=0}^{k-d} z^j .$$
2. Show that $\lambda(A_{G(d,k)})$ is the largest positive solution of the equation
$$z^{k+2} - z^{k+1} - z^{k-d+1} + 1 = 0 .$$
3. Show that for $d = 0$,
$$\lambda(A_{G(0,k)}) \le 2 - \frac{1}{2^{k+1}} .$$
4. Extending the definition of $G(d, k)$ to $k = \infty$, show that $\lambda(A_{G(d,\infty)})$ is the largest positive solution of the equation
$$z^{d+1} - z^d - 1 = 0 .$$
5. [AS87] Show that for $d \ge 1$,
$$\lambda(A_{G(d,\infty)}) = \lambda(A_{G(d-1,2d-1)}) .$$
6. [ForsB88] Show that for $d \ge 0$,
$$\lambda(A_{G(d,2d)}) = \lambda(A_{G(d+1,3d+1)}) .$$

Problem 3.20 [C70] For a nonnegative integer B, let G_B be the Shannon cover in Figure 1.14 of the B-charge constraint and let $\chi_B(z) = \det(zI - A_{G_B})$ be the characteristic polynomial of A_{G_B}.

1. Show that $\chi_0(z) = z$ and $\chi_1(z) = z^2 - 1$.
2. Show that for $B \ge 2$,
$$\chi_B(z) = z \cdot \chi_{B-1}(z) - \chi_{B-2}(z) .$$

3. Show that

$$\chi_B(2\cos x) = \frac{\sin{(B+2)x}}{\sin x}$$

and, so, for $|z| \le 2$,

$$\chi_B(z) = \frac{\sin\left((B+2)\cos^{-1}(z/2)\right)}{\sin\left(\cos^{-1}(z/2)\right)} \ .$$

Hint: Make use of the trigonometric identity

$$\sin{(Bx)} + \sin{(B+2)x} = 2\cos x \cdot \sin{(B+1)x} \ .$$

The polynomials $\chi_B(2z)$ are known as *Chebyshev polynomials of the second kind.* See [AbS65, pp. 774–776].

4. Show that the eigenvalues of A_{G_B} are given by

$$\lambda_i = 2\cos\left(\frac{\pi i}{B+2}\right) \ , \quad i = 1, 2, \ldots, B+1 \ .$$

5. Let the graph H_B be obtained from G_B by adding an edge from state 0 (the leftmost state in Figure 1.14) to state B (the rightmost state), and another edge from state B to state 0. Denote by $\overline{\chi}_B(z)$ the characteristic polynomial of H_B. Show that

$$\overline{\chi}_B(2\cos x) = 2\cos{(B+1)x} \ - \ 2$$

and, so, for $|z| \le 2$,

$$\overline{\chi}_B(z) = 2\cos\left((B+1)\cos^{-1}(z/2)\right) \ - \ 2 \ .$$

The polynomials $\overline{\chi}_B(2z) + 2$ are known as *Chebyshev polynomials of the first kind.*

6. Show that the eigenvalues of A_{H_B} are given by

$$\lambda_i = 2\cos\left(\frac{\pi i}{B+1}\right) \ , \quad i = 1, 2, \ldots, B+1 \ .$$

Problem 3.21 [Sen80] Let A be a nonnegative square matrix, not necessarily irreducible.

1. Show that there always exists a *nonnegative* real eigenvector associated with the largest eigenvalue $\lambda(A)$.

   Hint: Present A as a limit of an infinite sequence of irreducible matrices A_i. Show that the largest eigenvalues $\lambda(A_i)$ converge to $\lambda(A)$, and a respective eigenvector of A can be presented as a limit of eigenvectors of (a subsequence of) the A_i's.

2. Does there always exist such an eigenvector that is *strictly positive*?

Problem 3.22 Let A be a nonnegative irreducible matrix and let μ be an eigenvalue of A. Show that there exists a nonnegative real eigenvector associated with μ only if $\mu = \lambda(A)$.

Problem 3.23 Let H be the graph in Figure 2.26. Show that $\lambda(A_H) = 2$ by finding a strictly positive vector $\mathbf{x}$ such that $A_H\mathbf{x} = 2\mathbf{x}$.

Problem 3.24 Let A be a nonnegative irreducible matrix.

1. Show that $(A)_{u,u} \le \lambda(A)$ for every row index u. When does the inequality hold with equality?
2. Two rows in A indexed by u and u' are called *twin rows* if the following two conditions hold:
   (a) $(A)_{u,v} = (A)_{u',v}$ for every column index $v \notin \{u, u'\}$;
   (b) $(A)_{u,u} + (A)_{u,u'} = (A)_{u',u} + (A)_{u',u'}$.

   Let $\mathbf{x} = (x_v)_v$ be an eigenvector of A associated with an eigenvalue μ. Show that if u and u' index twin rows and $(A)_{u,u} - (A)_{u',u} \ne \mu$, then $x_u = x_{u'}$. Provide an example where $(A)_{u,u} - (A)_{u',u} = \mu$ and $x_u \ne x_{u'}$.
3. Show that if $\mathbf{x} = (x_v)_v$ is an eigenvector of A associated with $\lambda(A)$ and u and u' index twin rows, then $x_u = x_{u'}$.
4. Suppose that A is the adjacency matrix of a deterministic graph G and let the states u and u' index twin rows in A (note that the sets of outgoing edges of u and u' in G do not necessarily have the same sets of labels). Let the graph H be obtained from G by redirecting the incoming edges of u in G into u' and deleting state u with its outgoing edges. Show that H is irreducible and that $\mathsf{cap}(S(G)) = \mathsf{cap}(S(H))$.

Problem 3.25 Let A be a nonnegative integer irreducible matrix of order m and let $\mathbf{x} = (x_v)_v$ be a positive real eigenvector associated with $\lambda(A)$. Denote by $x_{\max}$ and $x_{\min}$ the largest and smallest entries in $\mathbf{x}$, respectively. Show that

$$\frac{x_{\max}}{x_{\min}} \le (\lambda(A))^{m-1} .$$

Hint: Think of A as the adjacency matrix of a graph G, and show that if there is an edge from u to v in G, then $x_v = \lambda(A)x_u$.

Problem 3.26 Let G be a nontrivial irreducible lossless graph and let G' be obtained from G by deleting an edge from G. Show that

$$\mathsf{cap}(S(G')) < \mathsf{cap}(S(G)) .$$

Hint: Use Proposition 3.12.

Problem 3.27 Let G be a nontrivial irreducible graph. It follows from Proposition 3.12 that if G' is obtained from G by deleting *any* edge from G, then $\lambda(A_{G'}) < \lambda(A_G)$. Show that there exists at least one edge in G, the deletion of which produces a graph G' for which

$$\lambda(A_G) - 1 \le \lambda(A_{G'}) \le \lambda(A_G) .$$

Problem 3.28 Let S_1 and S_2 be irreducible constrained systems with the same capacity. Show that if $S_1 \subseteq S_2$ then $S_1 = S_2$.

Hint: Assume to the contrary that there is a word in $S_2 \setminus S_1$ whose length, ℓ, is relatively prime to the period of the Shannon cover of S_2; then consider the constrained systems S_1^ℓ and S_2^ℓ.

Problem 3.29 Let S_0 be an irreducible constrained system and let G be a graph such that $S_0 \subseteq S(G)$ and $\mathsf{cap}(S_0) = \mathsf{cap}(S(G))$. Show that there is an irreducible component H of G such that $S_0 = S(H)$.

Problem 3.30 Let S be the constrained system over the alphabet $\Sigma = \{a, b, c, d, e, f, g\}$ generated by the graph G in Figure 2.22.

1. Obtain the matrices A_G and A_{G^2}.
2. Show that $\lambda(A_G) = 3$ and find a nonnegative integer right eigenvector and a nonnegative integer left eigenvector associated with the eigenvalue 3.
3. Compute the other eigenvalues of A_G.
4. Compute the eigenvalues of A_{G^2}.
5. Compute the capacity of S.

Problem 3.31 Let G be an irreducible *lossy* graph. Show that $\mathsf{cap}(S(G)) < \log \lambda(A_G)$.

Problem 3.32 (Graphs with extremal eigenvalues [LM95])

1. Among all irreducible graphs with m states, find an irreducible graph G_m for which $\lambda(A_{G_m})$ is minimal.
2. Among all *primitive* graphs with three states, find a primitive graph H_m for which $\lambda(A_{H_3})$ is minimal.
3. Find a primitive graph H_m with m states for which $\lambda(A_{H_m})$ is minimal.

Problem 3.33 Show by example that there are *nonstationary* Markov chains $\mathcal{P}$ on irreducible graphs G such that $\mathsf{H}(\mathcal{P}) > \log \lambda(A_G)$.

Problem 3.34 Let Q be an irreducible stochastic $|V| \times V|$ matrix and let $\boldsymbol{\pi}^\top = (\pi_u)_{u \in V}$ be the vector as in Proposition 3.19. Denote by p the period of Q and let $C_0, C_1, \ldots, C_{\mathsf{p}-1}$ be the equivalence classes of the congruence relation defined on the states of the support graph of Q. Assume that rows in Q that are indexed by C_r precede those that are indexed by C_{r+1}.

1. Show that for $r = 0, 1, \ldots, \mathsf{p}-1$,

$$\sum_{u \in C_r} \pi_u = 1/\mathsf{p} \ .$$

2. Denote by $\boldsymbol{\pi}_r$ the subvector of $\boldsymbol{\pi}$ that is indexed by C_r; that is, $\boldsymbol{\pi}_r = (\pi_u)_{u \in C_r}$. Also, let $\mathbf{1}_r$ be an all-one vector of length $|C_r|$. Define the $|V| \times |V|$ matrix Π_Q by

$$\Pi_Q = \frac{1}{\mathsf{p}} \begin{pmatrix} \mathbf{1}_0 \boldsymbol{\pi}_0^\top & & & 0 \\ & \mathbf{1}_1 \boldsymbol{\pi}_1^\top & & \\ & & \ddots & \\ 0 & & & \mathbf{1}_{\mathsf{p}-1} \boldsymbol{\pi}_{\mathsf{p}-1}^\top \end{pmatrix} .$$

Show that for every $t \geq 0$,

$$Q^{\mathsf{p}t} = \Pi_Q + E^{(t)} \ ,$$

where $E^{(t)}$ satisfies

$$\sum_{u,v \in V} |(E^{(t)})_{u,v}| \leq \beta \cdot \alpha^t$$

for some $0 \leq \alpha < 1$ and $\beta > 0$ and every $t \geq 0$.

Hint: Apply Theorems 3.10 and 3.17 and make use of the rate of convergence of the limit in Theorem 3.17, as stated in (3.6).

Problem 3.35 (Autocorrelation and power spectral density) Let $\mathcal{P}$ be an irreducible Markov chain with period p on a labeled graph $G = (V, E, L)$ where $L : E \to \mathbb{R}$. Denote by $C_0, C_1, \ldots, C_{\mathsf{p}-1}$ the equivalence classes of the congruence relation defined on the states of G, and for $r = 0, 1, \ldots, \mathsf{p}-1$, let $\overline{L}_r$ be the conditional expectation

$$\overline{L}_r = \mathsf{E}_{\mathcal{P}} \left\{ L(e) \,|\, \sigma(e) \in C_r \right\} \ ,$$

where $\sigma(e) = \sigma_G(e)$ is the initial state of the edge e in G. Define the random sequence

$$\mathbf{X} = X_{-\ell} X_{-\ell+1} \ldots X_0 X_1 \ldots X_\ell$$

on paths

$$e_{-\ell} e_{-\ell+1} e_0 e_1 \ldots e_\ell$$

of length $2\ell+1$ in G by

$$X_i = L(e_i) - \overline{L}_{r(e_i)} \ ,$$

where $r(e_i)$ is the index r such that $\sigma(e_i) \in C_r$. The *autocorrelation* of $\mathbf{X}$ is defined by

$$R_{\mathbf{X}}(t, i) = \mathsf{E}_{\mathcal{P}} \left\{ X_i X_{i+t} \right\}$$

for every $-\ell \leq i, i+t \leq \ell$.

Denote by Q the transition matrix associated with $\mathcal{P}$ and by $\boldsymbol{\pi}^\top$ the state probability vector of $\mathcal{P}$.

1. Let e_i and e_{i+t} be the ith and $(i+t)$th edge, respectively, along a random path on G. Show that for every $t > 0$,

$$\mathsf{Prob}\left\{e_i = e \text{ and } e_{i+t} = e'\right\} = \mathcal{P}(e) \cdot (Q^{t-1})_{\tau(e),\sigma(e')} \cdot q_{e'} \;,$$

where the probability is taken with respect to $\mathcal{P}$ and $q_{e'}$ is the conditional probability of the edge e'.

2. Let B be the $|V| \times |V|$ matrix whose entries are defined for every $u, v \in V$ by

$$(B)_{u,v} = \sum_{e \in E_u \,:\, \tau(e)=v} (L(e) - \overline{L}_{\tau(e)}) \cdot q_e \;.$$

Show that for every $t > 0$,

$$R_{\mathbf{X}}(t,i) = \boldsymbol{\pi}^\top B Q^{t-1} B \mathbf{1} \;.$$

In particular, $R_{\mathbf{X}}(t,i)$ does not depend on i (provided that $\ell \le i, i{+}t \le \ell$) and will therefore be denoted hereafter by $R_{\mathbf{X}}(t)$.

3. Let Π_Q be the matrix defined in part 2 of Problem 3.34. Show that $\boldsymbol{\pi}^\top B \Pi_Q = \Pi_Q B \mathbf{1} = 0$.

4. Show that there exist $0 \le \alpha < 1$ and $\beta > 0$ such that for every $t > 0$,

$$|R_{\mathbf{X}}(t)| \le \beta \cdot \alpha^{t-1} \;.$$

Hint: Apply Problem 3.34.

5. Show that the semi-infinite series

$$\sum_{t=1}^{\infty} R_{\mathbf{X}}(t) z^{-t}$$

converges for all complex values z with $|z| \ge 1$.

6. Show that

$$\sum_{t=1}^{\infty} R_{\mathbf{X}}(t) z^{-t} = \boldsymbol{\pi}^\top B (zI - Q)^{-1} B \mathbf{1}$$

for all $|z| \ge 1$, except when $z^{\mathsf{p}} = 1$ (in particular, verify that the matrix $zI - Q$ is nonsingular in this range of convergence). Extend the result also to the limit $z \to e^{\jmath 2\pi r/\mathsf{p}}$, where $\jmath = \sqrt{-1}$ and $r = 0, 1, \ldots, \mathsf{p}{-}1$.

7. The *power spectral density* of the random sequence $\mathbf{X}$ is defined as the two-sided Fourier transform of $R_{\mathbf{X}}(t)$; namely, it is the function $\Psi_{\mathbf{X}}(f) : [0,1] \to \mathbb{C}$ which is given by

$$\Psi_{\mathbf{X}}(f) = \sum_{t=-\infty}^{\infty} R_{\mathbf{X}}(t) e^{-\jmath 2\pi t f}$$

(note that $R_{\mathbf{X}}(-t) = R_{\mathbf{X}}(t)$ and that by part 5 the bi-infinite sum indeed converges). Show that for every $f \in [0,1] \setminus \{r/\mathsf{p}\}_{r=0}^{\mathsf{p}-1}$,

$$\Psi_{\mathbf{X}}(f) = R_{\mathbf{X}}(0) + 2 \cdot \mathrm{Re}\left\{\boldsymbol{\pi}^\top B (e^{\jmath 2\pi f} I - Q)^{-1} B \mathbf{1}\right\} \;,$$

where Re$\{\cdot\}$ stands for the real value and

$$R_{\mathbf{X}}(0) = \sum_{e \in E} \mathcal{P}(e)(L(e) - \overline{L}_{r(e)})^2 \ .$$

8. Let $f \mapsto \Phi_{\mathbf{X}}(f)$ be the (two-sided) Fourier transform of $\mathbf{X}$, namely,

$$\Phi_{\mathbf{X}}(f) = \sum_{i=-\ell}^{\ell} X_i e^{-\jmath 2\pi i f} \ .$$

Show that

$$\Psi_{\mathbf{X}}(f) = \lim_{\ell \to \infty} \frac{1}{2\ell+1} \mathsf{E}_{\mathcal{P}} \left\{ |\Phi_{\mathbf{X}}(f)|^2 \right\} \ .$$

9. Let G be the Shannon cover of the 2-charge constrained system, as shown in Figure 3.2. Consider the irreducible Markov chain $\mathcal{P}$ on G that is defined by the transition matrix

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} \ .$$

Show that $\mathcal{P}$ is a maxentropic Markov chain and that

$$\Psi_{\mathbf{X}}(f) = 1 - \cos 2\pi f \ .$$

Problem 3.36 The purpose of this problem is proving Theorem 3.21 for every irreducible Markov chain $\mathcal{P}$ on G. As with the case of independent random variables, the proof is based upon Chebyshev inequality, which states that for every random variable Y with zero expected value,

$$\mathsf{Prob}\left\{|Y| \ge \epsilon\right\} = \mathsf{Prob}\left\{\frac{Y^2}{\epsilon^2} \ge 1\right\} \le \mathsf{E}\left\{\frac{Y^2}{\epsilon^2}\right\} = \frac{\mathsf{E}\left\{Y^2\right\}}{\epsilon^2} \ .$$

Letting p be the period of G, we use the notations C_r and $\overline{L}_r$ as in Problem 3.35.

For a positive integer ℓ, define the random variable $Y_\ell(\gamma)$ on a path $\gamma = e_1 e_2 \ldots e_\ell$ by

$$Y_\ell(\gamma) = Y_\ell = \frac{1}{\ell} \sum_{i=1}^{\ell} X_i \ ,$$

where

$$X_i = L(e_i) - \overline{L}_{r(e)}$$

and $r(e)$ is the index r such that $\sigma_G(e) \in C_r$. Hereafter, all the expectations are taken with respect to $\mathcal{P}$.

1. Show that

$$\overline{L} = \frac{1}{\mathsf{p}} \sum_{r=0}^{\mathsf{p}-1} \overline{L}_r \ .$$

Hint: Apply part 1 of Problem 3.34.

2. Show that $\mathsf{E}\{Y_\ell\} = \mathsf{E}\{X_i\} = 0$.
3. Let Z_ℓ and $\overline{L}$ be defined as in the statement of Theorem 3.21. Show that $Y_\ell = Z_\ell - \overline{L}$.
4. Show that there is a real constant η such that
$$\mathsf{E}\left\{Y_\ell^2\right\} \le \frac{1}{\ell}\left(\mathsf{E}\left\{X_1^2\right\} + \eta\right) .$$
Hint: Write
$$\mathsf{E}\left\{\left(\sum_{i=1}^{\ell} X_i\right)^2\right\} = \sum_{i=1}^{\ell}\left(\mathsf{E}\left\{X_i^2\right\} + 2\sum_{t=1}^{\ell-i}\mathsf{E}\left\{X_i X_{i+t}\right\}\right) .$$
Then use part 4 in Problem 3.35.
5. Complete the proof of Theorem 3.21 by applying Chebyshev inequality to Y_ℓ.

Problem 3.37 Let $\mathcal{P}$ be an irreducible Markov chain on G. Show that for every $\delta > 0$ there exist ϵ and N such that every $(\mathcal{P}, \epsilon)$-strongly-typical path γ of length $\ell \ge N$ in G is $(\mathcal{P}, \delta)$-typical.

Hint: Let $\boldsymbol{\pi} = (\pi_u)_u$ be the stationary probability vector of $\mathcal{P}$ and let q_e denote the conditional probability of an edge e in G. Consider a $(\mathcal{P}, \epsilon)$-strongly-typical path of length ℓ in G that starts at state u_0. First argue that
$$\pi_{u_0} \prod_{e\,:\,q_e>0} q_e^{\ell(\pi_{\sigma(e)}q_e+\epsilon)} \le \mathcal{P}(\gamma) \le \pi_{u_0} \prod_{e\,:\,q_e>0} q_e^{\ell(\pi_{\sigma(e)}q_e-\epsilon)} .$$
Then deduce that
$$\left|\mathsf{H}(\mathcal{P}) + \frac{\log \mathcal{P}(\gamma)}{\ell} - \frac{\log \pi_{u_0}}{\ell}\right| \le -\epsilon \sum_{e\,:\,q_e>0} \log q_e .$$
Finally, given δ, pick ϵ and N so that
$$\delta \ge -\frac{\log\left(\min_u \pi_u\right)}{N} - \epsilon \sum_{e\,:\,q_e>0} \log q_e .$$

Problem 3.38 Show that Theorem 3.22 holds also when γ ranges over all $(\mathcal{P}, \epsilon)$-strongly-typical paths of length ℓ.

Hint: Apply Theorem 3.21 with $L : E \to \mathbb{R}$ being the indicator function $\mathcal{I}_e$ of an edge $e \in E$; i.e., $\mathcal{I}_e(e')$ takes the value 1 if $e' = e$ and is zero for $e' \in E \setminus \{e\}$.