I.2
Analytical methods

We shall now consider various ways to prove ergodicity of independent operators. The main topic of Chapters 3 and 4 is the general case: non-homogeneous operators on arbitrary graphs.

Chapter 3 treats the method first presented in [96] based on the estimation of dependence of an automaton's behaviour on its neighbours' states; here we give a modified version of this method, following [55].

Chapter 4 follows [97]; it presents a method that considers transitional operators as they work in the functional space $F(X)$.

Chapter 5 describes cluster expansions. Chapter 6 treats one-dimensional operators belonging to the class $\mathscr{P}_1$. Here we prove a sufficient condition of their ergodicity which seems to us to be well on the way to the necessary one.

In Chapter 7 we gain numerical concreteness at the price of generality. We consider a very narrow class of one-dimensional homogeneous operators with two neighbours but examine in detail the domain of parameters' values where ergodicity is provable by the methods we describe here.

Chapter 3
The coupling method

Let us have a space $X = \prod_{h \in V} X_h$ and a graph $\Gamma(V, \mathscr{U})$ of neighbours. To define an independent operator P it remains only to give transition probabilities. We denote them here as $\theta_z \in \mathscr{M}(X_h)$ where $z \in X_{U(h)}$. (Of course, θ_z depends on h, which is omitted.) The measures θ_z define operator P in the following way: for any $x \in X$

$$xP = \prod_{h \in V} \theta_{x_{U(h)}}.$$

There is a very simple case, in which every θ_z does not depend on its argument z at all. In this case, the measure xP does not depend on x, whence P is obviously ergodic. There are cases, which are close to this one, in which every θ_z 'weakly' depends on z; in this section we prove that in some of these cases P is ergodic too.

First, we explain the idea of this proof for a very special case, in which $\Gamma = \Gamma_1$, all X_h equal some X_0, and P is homogeneous. Suppose that at every point h the measure θ_z equals some mixture of two measures ξ and $\eta_z \in \mathscr{M}(X_0)$:

$$\theta_z = \varepsilon\xi + (1 - \varepsilon)\eta_z \tag{3.1}$$

where $0 \leqslant \varepsilon \leqslant 1$ and the measure $\xi \in \mathscr{M}(X_0)$ does not depend on z. We shall prove that $\varepsilon > \varepsilon^*$ makes P ergodic (with any ξ, η_z) where ε^* is the critical value of Example 1.2.

To this end, let us represent behaviour of P in the following way. Let x_i^t be the state of the point $i \in \mathbb{Z}$ at the time $t = 0, 1, 2, \ldots$, as if we were considering the evolution measure $\tilde{\mu}$ on the evolution graph. But now x_i^t has three states: 0, 1 and $*$ (which means uncertainty). Initially all the automata are in the uncertain state $*$. After that the uncertain states become substituted by zeros and ones according to the following sequence of rules:

1 At first, every value of x_i^t for all $i \in \mathbb{Z}$, $t > 0$ is changed to the value of the random variable ξ (0 or 1) with probability ε and remains uncertain with probability $1 - \varepsilon$.
2 Second, if x_i^t is yet uncertain, but both x_{i-1}^{t-1} and x_i^{t-1} are certain now, then x_i^t becomes equal the value of η_z where $z = (x_{i-1}^{t-1}, x_i^{t-1})$. This is repeated as long as there are points (i,t) where $x_i^t = *$ but $x_i^{t-1} \neq *$ and $x_{i-1}^{t-1} \neq *$.
3 Now we change the values of all x_i^0 (which have been uncertain till now) to zeros and ones arbitrarily, thus choosing the initial states.
4 Next, do the same as in rule 2.

These four rules result in all x_i^t becoming zeros and ones, and their distribution coincides with the measure $\tilde{\mu}$ generated by P and the initial measure (concentrated in the initial state). This is provable by induction. Now note that the result of rules 1 and 2 does not depend on the initial state, because it has got fixed when the initial state was yet uncertain. Hence, if the certain states (0 and 1) resulting from rules 1 and 2 prevail with $t \to \infty$, the operator P is ergodic.

But our rules 1 and 2 are such that the distribution of uncertain states depends only on ε (not on ξ or η_z) and equals the distribution of zeros in Example 1.2 with the initial state 'all zeros'.

Now let us have an operator and want to prove its ergodicity. For that we present its transition probabilities in the form of rule 1 above, and try to make ε as large as possible. One can see that for a given P the largest possible value of ε in rule 1 equals

$$\varepsilon_0 = \sum_{y \in X_0} \min_{z \in X_{U(h)}} \theta_z^y,$$

where θ_z^y are the transition probabilities of P. If ε_0 happens to exceed ε^* of Example 1.2, ergodicity of P is proved. This argument is generalisable to treat non-homogeneous but independent operators. We shall do that later, but for the present, following [96,55], proceed in other terms, less explanative but formally more coherent. And we shall obtain a better estimation.

Definition 3.1 Let μ_1 and μ_2 be measures on spaces X_1 and X_2 respectively. A measure μ on the product space $X_1 \times X_2$ is termed a coupling of μ_1 and μ_2 if $\mu(C_1 \times X_2) = \mu_1(C_1)$ and $\mu(X_1 \times C_2) = \mu_2(C_2)$ for any measurable $C_1 \subset X_1$ and $C_2 \subset X_2$. Of course, for most given X_1, X_2, μ_1 and μ_2 there are many couplings, including, say, the product measure $\mu_1 \times \mu_2$. But the following special coupling is particularly important for us. First assume that X_1 and X_2 are equal and finite:

$$X_1 = X_2 = X \quad \text{and} \quad |X| < \infty.$$

The coupling we need has the largest possible sum of its values at the diagonal:

$$\sum_{x \in X} \mu(x,x).$$

It can be defined explicitly in the following way. First we define it at the diagonal:

$$\mu(x,x) = \min(\mu_1(x), \mu_2(x))$$

for all $x \in X$. Then we define it at the rest of $X \times X$ by setting for all $x \neq y$

$$\mu(x,y) = \frac{1}{\nu}(\mu_1(x) - \mu(x,x))(\mu_2(y) - \mu(y,y)),$$

where

$$\nu = \frac{1}{2}\mathrm{Var}(\mu_1 - \mu_2) = \frac{1}{2}\sum_{x \in X} |\mu_1(x) - \mu_2(x)| =$$
$$= \sum_{x \in X} (\mu_1(x) - \mu(x,x)) = \sum_{x \in X} (\mu_2(x) - \mu(x,x)).$$

Denote this coupling $\mu_1 \odot \mu_2$. The sum of its values at the diagonal equals

$$\sum_{x \in X} \mu_1 \odot \mu_2(x,x) = 1 - \frac{1}{2}\mathrm{Var}(\mu_1 - \mu_2).$$

If $\mu_1 = \mu_2$, the whole of our $\mu_1 \odot \mu_2$ coupling is concentrated at the diagonal. Informally but importantly, if $\mu_1 \approx \mu_2$, the bulk of $\mu_1 \odot \mu_2$ is at the diagonal.

Now let us proceed to operators. Let operator P_1 act from X_1 to X_1 and P_2 act from X_2 to X_2.

Definition 3.2 An operator Φ from $X_1 \times X_2$ to $X_1 \times X_2$ is termed a coupling of P_1 and P_2 if either of the two following equivalent assertions holds:

1. For every $x = (x_1,x_2) \in X_1 \times X_2$ the measure $x\Phi$ is a coupling of x_1P_1 and x_2P_2.
2. For every function $\varphi \in F(X_1 \times X_2)$ which actually depends only on one of its two arguments the following holds:

$$\text{if} \quad \varphi \in F(X_1) \quad \text{then} \quad \Phi\varphi = P_1\varphi,$$
$$\text{if} \quad \varphi \in F(X_2) \quad \text{then} \quad \Phi\varphi = P_2\varphi.$$

Equivalence of assertions 1 and 2 can be checked immediately. Note that assertion 2 implies that Φ^t is a coupling of P_1^t and P_2^t for any natural t.

Lemma 3.3 [55]. Let an operator P act from X to X. Suppose that there is such a coupling Φ of P with P that for any $\psi \in F(X,X)$ the condition

$$\psi(x,x) = 0 \quad \text{for all} \quad x$$

implies

$$\lim_{t\to\infty} \Phi^t\psi = 0.$$

Then P has only one invariant measure. If the same condition implies uniform convergence of $\Phi^t\psi$ to 0, then P is ergodic.

Proof. Take any $\varphi \in F(X)$ and put $\psi(x,y) = \varphi(x) - \varphi(y)$. The lemma's condition provides that for any x, y the difference

$$P^t\varphi(x) - P^t\varphi(y) = \Phi^t\psi(x,y)$$

tends to 0 with $t \to \infty$. If μ is invariant for P, the difference

$$P^t\varphi(x) - \mu\varphi = \int_X (P^t\varphi(x) - P^t\varphi(y))d\mu(y)$$

tends to 0 for any x. Hence, if ν is invariant for P too, $\mu\varphi = \nu\varphi$. But we have taken φ arbitrarily, whence $\mu = \nu$, and we have proved the uniqueness of invariant measure.

Now suppose that $\Phi^t\psi$ uniformly tends to 0. Then the difference

$$P^t\varphi(x) - \mu\varphi$$

tends to 0 uniformly too and P is ergodic according to Definition 2.7.

Now, to formulate a theorem, we return to the beginning of the section: an independent operator P defined by parameters θ_z for all $h \in V$, $z \in X_{U(h)}$. For the given P we define another operator $Q \in \mathcal{P}_1(\{0,1\}^V)$ defined by parameters

$$\theta_z = 1 - \tfrac{1}{2}\max \operatorname{Var}(\theta_v - \theta_w) \tag{3.2}$$

where the maximum is taken over such v, $w \in X_{U(h)}$ for which

$$k \in U(h),\ z_k = 1 \Rightarrow v_k = w_k.$$

This Q is termed the minorant of P. It is monotone (in the sense $0 \prec 1$).

Theorem 3.4 If the minorant of P is ergodic, P is ergodic too.

Proof. We use the notations

$$X = \prod_{h \in V} X_h,\ Y = \{0,1\}^V.$$

Of course, P acts from X to X and Q acts from Y to Y. We define a mapping D: $X \times X \to Y$ by the following rule:

$$D(x',x'') = y$$

where

$$y_h = \begin{cases} 1 & \text{if} \quad x'_h = x''_h, \\ 0 & \text{if} \quad x'_h \neq x''_h. \end{cases}$$

for all $h \in V$. According to our convention, the same letter D stands for the corresponding mapping from $F(Y)$ to $F(X \times X)$.

Now we define an independent operator Φ from $X \times X$ to $X \times X$ which transforms any point (x',x'') into the product measure $\prod_{h \in V} \xi_h$, every factor ξ_h of which is the special coupling defined as above:

$$\xi_h = \theta_{z'} \odot \theta_{z''}$$

where z' and z'' are projections of x' and x'' to $X_{U(h)}$. Evidently, Φ is a coupling of P with P. Based on Corollary 2.13, it is easy to prove that $DQ \prec \Phi D$.

Now we suppose that Q is ergodic. This means that for any $\varphi \in F(Y)$

$$\lim Q^t\varphi(y) = \varphi(\mathbf{1}) \tag{3.3}$$

uniformly in $y \in Y$, where $\mathbf{1}$ stands for the element 'all ones' of Y. Let us prove that Φ satisfies the condition of Lemma 3.3. Choose some $\psi \in F(X \times X)$ such that $\psi(x,x) = 0$ for any $x \in X$. Denote

$$\varphi_1(y) = \min \psi(x',x''),$$
$$\varphi_2(y) = \max \psi(x',x''),$$

where the minimum and the maximum are taken over all (x',x'') such that $y \prec D(x',x'')$. Both φ_1 and φ_2 are monotone and

$$\varphi_1(\mathbf{1}) = \varphi_2(\mathbf{1}) = 0,\ D\varphi_1 \leqslant \psi \leqslant D\varphi_2.$$

Using Proposition 2.14, this implies

$$DQ^t\varphi_1 \leqslant \Phi^t D\varphi_1 \leqslant \Phi^t\psi \leqslant \Phi^t D\varphi_2 \leqslant DQ^t\varphi_2. \tag{3.4}$$

Application of Lemma 3.3 to the leftmost and rightmost terms of (3.4) proves that $\Phi^t\psi$ tends to 0 with $t \to \infty$ uniformly in $(x', x'') \in X \times X$. Now Lemma 3.3 proves our theorem.

In this proof the locality of P is not essential.

In [96], where coupling was introduced and Lemma 3.3 was proved, the same method was used to prove ergodicity in the more general case:

every automaton might have infinitely many neighbours but had to depend on them weakly enough.

But the independence of P is essential in this proof as it enables us to construct Q. Let P be a homogeneous operator on the graph Γ_1 with $X = \{0;1\}$. It has eight transition probabilities, of which four are independent; denote the probabilities $\theta(1|x_{U(h)})$ by θ_{00}, θ_{01}, θ_{10}, θ_{11} where the two indices stand for x^t_{h-1} and x^t_h. Define an operator $Q \in \mathscr{P}_1$ with the corresponding probabilities:

$$\bar{\theta}_{00} = 1 - \max_{x_1,x_2,y_1,y_2} |\theta_{x_1x_2} - \theta_{y_1y_2}|,$$

$$\begin{aligned}\bar{\theta}_{01} &= 1 - \max\{|\theta_{01}-\theta_{11}|,|\theta_{00}-\theta_{10}|\},\\ \bar{\theta}_{10} &= 1 - \max\{|\theta_{10}-\theta_{11}|,|\theta_{00}-\theta_{01}|\},\\ \bar{\theta}_{11} &= 1.\end{aligned}$$

Corollary 3.5 Ergodicity of Q implies ergodicity of P. To prove it note that

$$\mathrm{Var}(\eta - \zeta) = |\eta(1) - \zeta(1)|$$

for any two measures η, ζ on the set $\{0;1\}$. Hence θ_z are defined according to (3.2).

When applying Theorem 3.4 and Corollary 3.5, Proposition 2.17 is useful too. Let us show it in the homogeneous case. Put

$$\varkappa_i = 1 - \theta_z$$

for all $i = 1, \ldots, R$ where $z \in Y_{U(h)}$ has the ith component equal to 0 and all the other components equal to 1. Evidently, this satisfies all the conditions of Proposition 2.17 except, possibly, the inequality $\Sigma\varkappa_i < 1$. Thus, assuming this inequality to be true too, is implied ergodicity of Q.

For example, consider the majority voting operator of Examples 1.1a and 1.1b. The corresponding Q has parameters

$$\theta_z = 1 - |1 - 2\varepsilon|$$

for all $z \neq \mathbf{1}$. We know that ergodicity of Q (and P) is guaranteed by $|1 - 2\varepsilon| < \frac{1}{3}$, that is, by

$$\tfrac{1}{3} < \varepsilon < \tfrac{2}{3}.$$

Various applications of the 'coupling' method to continuous time systems are exposed thoroughly in [48]. In particular, this method may be used to prove that the set of invariant measures of the monotone (i.e. attractive) process on $\{0,1\}^{\mathbb{Z}}$ is the segment $[\nu_0,\nu_1]$, where ν_0 and ν_1 are the limit points of the process for initial states $\delta_\mathbf{0}$ and $\delta_\mathbf{1}$ respectively.

Another method of proving ergodicity in the discrete time case has been developed in [6–8]. It is based on estimation of 'lost information' and is suitable for systems where automata weakly depend on each other. The next chapter presents still another method of proving ergodicity, which is especially good for numerical estimations.

Chapter 4
Characteristic polynomials

This chapter describes in a general way an algebraic method, which has been illustrated by Example 1.3. We use it to prove ergodicity of some operators. However, it may be useful in answering some other questions, including investigation of the set of invariant measures. In some cases this method almost coincides with the dual process method; this interpretation allows us to apply it to the continuous time case (see [31,35,48]).

For any $\varphi \in F(X)$ denote by $\omega(\varphi) = \sup_{x,y \in X} |\varphi(x) - \varphi(y)|$, the range of φ. Remember that a semi-norm stands for a non-negative homogeneous function which satisfies the triangle inequality.

Definition 4.1 A ω-semi-norm is such a semi-norm $\|.\|$ in $F(x)$ that $\|\varphi\| \geqslant \omega(\varphi)$.

Let P stand for a local operator from X to X.

Proposition 4.2 If P is contractive with respect to some ω-semi-norm in $F(X)$, then P is ergodic.

Proof. For any $\varphi \in F(x)$ the sequence of ranges of P^t contracts:

$$[\inf P^{t+1}\varphi;\ \sup P^{t+1}\varphi] \subseteq [\inf P^t\varphi;\ \sup P^t\varphi].$$

The assumption that P is contractive implies $\|P^t\varphi\| \to 0$ whence $\omega(P^t\varphi) \to 0$ and $P^t\varphi$ uniformly converges to a constant. Thus, P is ergodic according to Definition 2.7.

Proposition 4.3 An operator P is given. Suppose that there is such a number $\varkappa < 1$ and such a subset $F_1 \subseteq F$ that F is the linear hull of F_1 and for any $\varphi \in F$ its image $P\varphi$ is presentable as such a finite linear combination

$$P\varphi = \sum_{r=1}^{R} \alpha_r \varphi_r \quad \text{where} \quad \varphi_r \in F_1, \tag{4.1}$$

that

$$\sum_{r=1}^{R} |\alpha_r| \omega(\varphi_r) \leqslant \varkappa \omega(\varphi). \tag{4.2}$$

Then P is ergodic.

Proof. For any $\varphi \in F$ put

$$\|\varphi\| = \inf \sum_{r=1}^{R} |\alpha_r| \omega(\varphi_r)$$

where the infimum is taken over all presentations $\varphi = \sum_{r=1}^{R} \alpha_r \varphi_r$ (where R is natural and all φ_r belong to F_1). It is easy to prove that $\|.\|$ thus defined is an ω-semi-norm in F and that P contracts it with the coefficient κ.

A similar assertion is proved in [47]. The following is its slight modification.

Proposition 4.4 Let $F_1 \subseteq F(X)$, the linear hull of F_1 be equal F, and $|\varphi(x)| \leqslant 1$ for all $\varphi \in F_1$, $x \in X$. Let there be such $\varkappa < 1$ that for any $\varphi \in F_1$ there is a presentation

$$P\varphi = \sum_{r=1}^{R} \alpha_r \varphi_r$$

where all $\varphi_r \in F_1$ and $\sum_{r=1}^{R} |\alpha_r| \leqslant \varkappa$. Then P is ergodic.

Its proof goes on the same lines. However, Propositions 4.3 and 4.4 are logically independent.

For the rest of this chapter we shall take $X = \{0;1\}^V$.

Let us denote for any $K \subset V$

$$\chi_K(x) = \begin{cases} 1 & \text{if} \quad x_h = 1 \quad \text{for all} \quad h \in K, \\ 0 & \text{otherwise}, \end{cases}$$

and term it the characteristic function of K. Of course, $\chi_\varnothing(x) \equiv 1$.

In these terms independence of an operator P from X to X means:

$$J \cap K = \varnothing \Rightarrow P(\chi_J \chi_K) = (P\chi_J)(P\chi_K).$$

Hence $P\chi_J = \prod_{h \in J} P\chi_h$.

Every $P\chi_h$ can be presented as a finite linear combination of some characteristic functions

$$P\chi_h = \sum \alpha_{hK}\chi_K \tag{4.3}$$

which will be termed the characteristic polynomial of P in the point h.

Knowing the characteristic polynomial, we can represent any $P\chi_K$ as a polynomial too by opening the brackets in

$$P\chi_K = \prod_{h\in K} \sum_J \alpha_{hJ}\chi_J. \tag{4.4}$$

Since any $\varphi \in F$ can be presented as a linear combination of characteristic functions, $P\varphi$ is presentable in an analogous way:

$$P\varphi = \sum_K \varphi_K(P\chi_K) = \sum \varphi_K \prod_{h\in K} \sum_J \alpha_{hJ}\chi_J. \tag{4.5}$$

Theorem 4.5 [47]. If

$$\sup_{h\in V} \sum_K |\alpha_{hK}| = \varkappa < 1, \tag{4.6}$$

operator P is ergodic.

Proof. Let F_1 be the totality of all χ_K (including $\chi_\varnothing \equiv 1$). It is sufficient to prove that our F_1 fits all the conditions of Proposition 4.3. Of course, the linear hull of F_1 is F. Now, after opening the brackets in (4.4) for any $K \neq \varnothing$ we obtain some linear combination with coefficients, the sum of modulos of which does not exceed $\varkappa$.

Thus, for any $K \neq \varnothing$ we have (4.1) satisfying (4.2). For $K = \varnothing$ (4.1) and (4.2) are evident, which ends the proof.

Functions which have other values than 0 or 1 are useful also to prove ergodicity of some other operators. Let us denote

$$\chi'_h = \begin{cases} a & \text{if} \quad x_h = 0, \\ b & \text{if} \quad x_h = 1 \end{cases}$$

where $a \neq b$ and $\chi'_\varnothing \equiv 1$, $\chi'_K = \prod_{h\in K} \chi'_h$ for $K \neq \varnothing$. Consider the characteristic polynomial of an independent operator P again: $P\chi'_h = \sum_K \alpha'_{hK}\chi'_K$.

Theorem 4.6 (A. M. Leontovitch). Let Γ be non-homogeneous but

$$\sup_{h\in V} |U(h)| = R < \infty.$$

Denote

$$\rho_m = |c_m| + |d_m|$$

where c_m and d_m are coefficients in the presentation

$$(\chi_h')^m = c_m + d_m\chi_h'.$$

Denote

$$\delta = \max\{|a|,|b|,1\},$$

$$\gamma = \max_{1\leq m\leq R}\{\delta^{R/m-1},\rho_m^{R/m}\}.$$

If

$$\sup_{h\in V}\gamma\sum_k|\alpha_{hK}'| = \varkappa < 1 \tag{4.7}$$

then P is ergodic.

The proof is based on the fact that P contracts with the coefficient $\varkappa$ the ω-semi-norm defined in the following way:

$$\|\chi_\varnothing'\| = 0,\ \|\chi_K'\| = 2\delta^{|K|}\quad\text{for}\quad K\neq\varnothing,$$

$$\left\|\sum_K\beta_K\chi_K'\right\| = \sum_{K\neq\varnothing}|\beta_K|\,\|\chi_K'\| = 2\sum_{K\neq\varnothing}|\beta_K|\sigma^{|K|}.$$

Let us apply Theorem 4.6 to the majority voting systems of Examples 1.1a and 1.1b. Take $a = -1$, $b = 1$. Then the characteristic polynomial is

$$P\chi_h' = (\tfrac{1}{2} - \varepsilon)(\chi_{h-1}' + \chi_h' + x_{h+1}' - \chi_{\{h-1,h,h+1\}}').$$

In this case $\delta = \rho_m = \gamma = 1$. The inequality (4.7) becomes

$$2|1 - 2\varepsilon| < 1\quad\text{or}\quad\tfrac{1}{4} < \varepsilon < \tfrac{3}{4}.$$

Thus, ergodicity of Examples 1.1a and 1.1b is proved in this range.

The characteristic functions are useful in another way: they help to examine how the invariant measure (which is unique in the treatable cases) of an operator depends on its parameters.

Proposition 4.7 In the conditions and notations of Theorem 4.5, in the domain

$$\sum_K|\alpha_K| < 1 \tag{4.8}$$

the values of the P operator's invariant measure analytically depend on the transition probabilities.

Proof. Note that in the domain (4.8) P has only one invariant measure according to Theorem 4.5. Denoting this sole invariant measure by μ_P we have

$$\mu_P\chi_K = \lim_{t\to\infty}(\mathbf{0}P^t)\chi_K = \lim_{t\to\infty}\varphi^t(\mathbf{0}) = \varphi^\infty(\mathbf{0})$$

where φ^t stands for $P^t\chi_K$. Now let us expand the formula (4.4) to the space of polynomials with complex coefficients. Let α_K run over all complex values satisfying $\sum_K |\alpha_K| < 1$. This domain is the union of open domains

$$\sum_K |\alpha_K| < \varkappa \tag{4.9}$$

over all $\varkappa < 1$. When proving Theorem 4.5 we in fact proved that in the domain (4.9) the operator P contracts with the coefficient $\varkappa$ the following ω-semi-norm:

$$\left\| \sum_K \varphi_K \chi_K \right\| = \sum_{K \neq \varnothing} |\varphi_K|.$$

But the same holds in the complex case too. Hence in the domain (4.9)

$$\omega(\varphi^t - \varphi^\infty) = \|\varphi^t - \varphi^\infty\| \leqslant \sum_{s=t}^{\infty} \|\varphi^s - \varphi^{s+1}\| \leqslant c \sum_{s=t}^{\infty} \varkappa^s.$$

Thus, (4.9) assumed, φ^t converges to φ^∞ uniformly in x and the value of $\varphi^t(\mathbf{0})$ for any t is a polynomial of α_K. So, according to the Weierstrass theorem $\varphi^\infty(\mathbf{0})$ depends on α_K analytically in any domain (4.9) and consequently in the domain (4.8) too. We have proved that (4.8) assumed $\mu_P \varphi$ depends on α_K analytically for any $\varphi \in F$. It remains to note that all α_K are linear combinations of transition probabilities. Proposition 4.7 is proved.

In the same way analyticity of all $\mu_P \chi'_K$ can be proved assuming that (4.7) holds for any homogeneous operator. These results can be expanded to non-homogeneous operators, understanding by analyticity of a function of a countable set of arguments its representability as an absolutely convergent series.

To conclude this chapter, we mention some elaborate estimations for the dimension of the set of invariant measures of a homogeneous operator $P \in \mathscr{P}_1$. Let $\Gamma(V, \mathscr{U})$ be a uniform graph with a transitive group G of automorphisms and $X = \{0;1\}^V$.

Definition 4.8 Term a measure $\mu \in \mathscr{M}(X)$ regular if

$$\lim_{t \to \infty} \sup_{h \in V} \mu \chi_{U^t(h)} = 0.$$

Proposition 4.9 [47,55]. Let $P \in \mathscr{P}_1$ be homogeneous and all the coefficients of its characteristic polynomial be non-negative. Let Γ be G-connected, which means that for any finite $K \subset V$ and $h \in V$ there are such $g \in G$ and $t_0 > 0$ that $g(K) \subset U^t(h)$ for all $t > t_0$. Then:

1 For any regular measure μ the sequence μP^t tends to $\mu_P = \lim_{t\to\infty} \mathbf{0}P^t$.

2 Any homogeneous invariant measure μ is the mixture of μ_P and $\delta_\mathbf{1}$:

$$\mu = \alpha\mu_P + (1 - \alpha)\delta_\mathbf{1} \quad \text{where} \quad 0 \leqslant \alpha \leqslant 1.$$

3 For any homogeneous measure μ the sequence μP^t converges.

4 If $\mu_P \neq \delta_\mathbf{1}$ then μ_p is regular and moreover $\lim_{|K|\to\infty} \mu_P\chi_K = 0$.

The condition of G-connectedness is essential because without it the totality of invariant measures can have dimension more than 1, as demonstrated by the following example. Let $V = \mathbb{Z}$, $U(h) = \{h-2,h\}$, $\theta_{11} = 1$, $\theta_{00} = \theta_{01} = \theta_{10} = \varepsilon$. This means that automata at even places and automata at odd places form two separated systems not interacting with each other. The behaviour of each system is the same as in Example 1.2. Thus each system with small values of ε has at least two mutually singular invariant measures μ_0 and $\mu_1 = \delta_\mathbf{1}$ whence P has at least three mutually singular homogeneous invariant measures:

$$\mu_0 \times \mu_0,\ \tfrac{1}{2}(\mu_0 \times \mu_1 + \mu_1 \times \mu_0),\ \mu_1 \times \mu_1.$$

Thus the dimension of the totality of homogeneous invariant measures of P is not less than 2. The same holds for

$$U(h) = \{h-1;h+1\}$$

which gives a connected but not G-connected graph.

There are analogous results for systems with continuous time [48].

Let us explain the similarity between the characteristic polynomials method and the duality method first introduced by Harris [35]. Let us restrict our attention to Example 1.2 and take the functions $\chi_i(x) = x_i$ and their products as the basis functions. The characteristic polynomial is

$$P\chi_i(x) = \varepsilon + (1 - \varepsilon)\chi_i\chi_{i+1}. \tag{4.10}$$

Its iterations result in formulae

$$P^t\chi_I = \sum_K \alpha_{I,K}^{(t)}\chi_K \tag{4.11}$$

for any finite $I \subset \mathbb{Z}$, where all K are finite subsets of $\mathbb{Z}$ too. The $\alpha_{I,K}^{(t)}$ are some positive coefficients (polynomials of ε and $(1 - \varepsilon)$ with positive coefficients) subject to $\sum_K \alpha_{I,K}^{(t)} = 1$. In particular, with $t = 1$

$$P\chi_I = \Sigma\alpha_{I,K}^{(1)}\chi_K = \prod_{i\in I}(\varepsilon + (1 - \varepsilon)\chi_i\chi_{i+1}).$$

Let us think of the numbers $\alpha_{I,K}^{(1)}$ as probabilities of transitions of a Markov chain η on the set S of all finite subsets of $\mathbb{Z}$; this chain goes from any $I \in S$ to any $K \in S$ with probability $\alpha_{I,K}^{(1)}$. Then $\alpha_{I,K}^{(t)}$ is this chain's probability of passing from I to K in t steps of time. Let η^t be the chain's state at time t, the initial state being η^0. Now we may consider the initial Markov process ξ (having the transition operator P) as having the same set S of states by representing any $x \in X$ by $\xi(x) = \{j \in \mathbb{Z}; x_j = 0\}$. Now we may interpret (4.11) in the probabilistic sense, which yields

$$\mathbb{P}(\xi^t \cap \eta^0 = \varnothing) = \mathbb{P}(\eta^t \cap \xi^0 = \varnothing). \tag{4.12}$$

This wonderful equality warrants the term 'duality': the processes η and ξ are dual. Note that these two processes have axes of time directed in opposite directions. In this particular example this opposition of times can be expressed in the percolation terms (see Example 1.2 and Chapter 8). One corollary of (4.12) is the following: for any x^0 having ξ^0 finite:

$$\lim_{t \to \infty} x^{\text{opt}} = \alpha\mu_\varepsilon + (1 - \alpha)\delta_1$$

where $\alpha = \mu_\varepsilon(x\colon x_i \geqslant \max\{x_i^0, x_{i-1}^0\})$.

The duality method is very useful in the voter model on the lattices $\mathbb{Z}^d$ too; this is the operator on $\{0;1\}^{\mathbb{Z}^d}$ having the linear characteristic polynomial of the form

$$P\chi_h = \Sigma\alpha_k\chi_{h+k}, \ \alpha_k > 0$$

It has δ_0 and δ_1 as invariant measures; in the cases $d = 1$ and $d = 2$ their convex hull exhausts the set of homogeneous invariant measures, but with any $d \geqslant 3$ the set has a continuum of different extreme points. This difference can be explained as follows: the dual chain is a system of coalescing random walks which are recurrent with $d = 1, 2$ but not with $d \geqslant 3$.

The notion of duality can be generalised so that to help to investigate many systems with continuous time just as the characteristic polynomials help to investigate systems with discrete time. These include the voter model investigated in the continuous time case especially well [3,5,39,48]; the self-dual contact process which is the continuous time analogue of the (4.10) operator; and many others (see [29,31,48]).

Chapter 5
Cluster expansions

This chapter is about some more ways (borrowed from statistical physics) of proving ergodicity of independent operators. As before, the transitional probabilities

$$\theta(x_h^t | x_{U(h)}^{t-1})$$

have to depend weakly on $x_{U(h)}^{t-1}$: $\theta(x_h^t | x_{U(h)}^{t-1}) \approx \eta(x_h^t)$. The present method also allows us to present the invariant measure parameters as power series in differences:

$$\theta(x_h^t | x_{U(h)}^{t-1}) - \eta(x_h^t) = \beta(x_h^t, x_{U(h)}^{t-1}). \tag{5.1}$$

As a preliminary illustration, let us apply this method to a finite Markov chain having n states 1, 2, ..., n and positive transition probabilities presented in a way analogous to (5.1):

$$p(i|j) = \eta(i) + \beta(i,j) \tag{5.2}$$

where

$$\sum_{i=1}^{n} \eta(i) = 1, \ \eta(i) \geqslant 0 \tag{5.3}$$

and all $\beta(i,j)$ are small. Clearly,

$$\sum_i \beta(i,j) = 0. \tag{5.4}$$

Let x_t stand for the state at time t. The conditional probability to get from one state to another in t steps is

$$\mathbb{P}(x_t = i_t | x_0 = i_0) = \sum_{i_1, \ldots, i_{t-1}} \prod_{k=1}^{t} P(i_k | i_{k-1}). \tag{5.5}$$

We want to substitute (5.2) into (5.5). This complicates the formula. To simplify it, we denote

$$\beta_k(i,j) = \sum_{j_1,\ldots,j_{k-1}} \beta(i,j_1)\beta(j_1,j_2)\ \ldots\ \beta(j_{k-1},j),$$

collect together members having k first factors of the same sort and use (5.3) and (5.4). This results in

$$\mathbb{P}(x_t = i|x_0 = i_0) = \eta(i) + \sum_{k=1}^{t-1}\sum_{j=1}^{n}\beta_k(i,j)\eta(j) + \beta_t(i,i_0) \qquad (5.6)$$

where only the last term depends on the initial state i_0. We have said informally that β are small. Let us formalise this as follows:

$$\forall_j\colon \sum_i |\beta(i,j)| \leqslant \alpha < 1.$$

Then

$$|\beta_k(i,j)| \leqslant \alpha^k$$

and (5.6) in the limit $t \to \infty$ presents the invariant measure values as absolutely convergent power series:

$$\begin{aligned}\mu(i) &= \eta(i) + \sum_{k=1}^{\infty}\sum_{j=1}^{n}\beta_k(i,j)\eta(j) = \\ &= \eta(i) + \sum_{k=1}^{\infty} G_k(i)\end{aligned} \qquad (5.7)$$

where

$$G_k(i) = \sum_{j=1}^{n}\beta_k(i,j)\eta(j)$$

Preliminaries over, let us now do the same for our operators. The method works for a variety of graphs including those having $V = \mathbb{Z}^d$. But we assume for simplicity $V = \mathbb{Z}$. We shall treat evolution measures $\tilde{\mu}$ on evolution graphs having $\tilde{V} = \mathbb{Z}^2$ as the set of vertices.

Let us term the set

$$B_{it} = \{(i,t)\} \cup \{(j,t-1), j \in U(i)\}$$

the 'branch' with the root $(i,t) \in \mathbb{Z}^2$. The points $(j,t-1)$, $j \in U(i)$ are termed 'ends' of this branch.

Definition 5.1

(1) A branch B_{jt-1} is termed *following* a branch B_{it} if $(j,t-1)$ is an end of B_{it}.

(2) A finite set of branches is termed a *crown* with one root (i,t) if it includes B_{it} and every one of its other branches follows another of its branches.

The union of points of branches of a crown C is termed crown C also. A single point (i,t) is also a crown with the root (i,t).

(3) A finite set of points is termed a crown with roots (i_1,t), (i_2,t), ..., (i_k,t) if it is a union of k crowns with these roots.

To every state $x_C \in X_C$ of a crown C we ascribe some value $\gamma(x_C)$ based on denotations of (5.1) by the following formula:

$$\gamma(x_C) = \Pi\beta(x_h^t, x_{U(h)}^{t-1}) \cdot \Pi\eta(x_h^t)$$

where the left-hand product is taken over all roots (h,t) of branches in C and the right-hand product is taken over all free ends (h,t) of C.

Lemma 5.2 Suppose that the system's initial measure is independent and is the product of measures equal to η in every point $(i,0)$. Then the probability

$$\tilde{\mu}(x_i^t = a)$$

equals the sum of $\gamma(x_C)$ over all $x_C \in X_C$ such that $x_i^t = a$ where C runs over all crowns having one root (i,t) in positive-time half-plane.

Proof. Just from definitions, $\tilde{\mu}(x_i^t = a)$ equals

$$\sum \left(\prod_{s=0}^{t-1} \prod_{j=i-rs}^{i+rs} \mathbb{P}(x_j^{t-s} | x^{t-s-1}) \right) \prod_{j=i-rt}^{i+rt} \eta(x_j^0) \tag{5.8}$$

over all x_j^u where j and u run over the given ranges with the condition $x_i^t = a$. Substituting (5.1) here and opening the brackets we prove the lemma.

Theorem 5.3 For any n, R there is such $\alpha = \alpha(n,R) > 0$ that the following holds. Let P be an independent homogeneous operator on $X = X_0^V$ where

$$|X_0| = n,\ V = \mathbb{Z},\ U(h) = \{j\colon |j-h| \leq r\}.$$

Assume (5.1) where all

$$|\beta(.,.)| \leq \alpha.$$

Then P is ergodic and any value of its invariant measure μ_P:

$$\mu_P(x_{i_1}=a_1, \ldots, x_{i_k}=a_k)$$

equals the sum of $\gamma(x_C)$ over all $x_C \in X_C$ such that

$$x_{i_1}^t = a_1, \ \ldots, \ x_{i_k}^t = a_k$$

where C runs over all crowns with roots (i_1,t), ..., (i_k,t). Here the value of t is arbitrary.

Proof. Denote $|U(h)| = R$ and remember that $|X_0| = n$. Note that:

1 The number of crowns with one root (i,t) consisting of k branches does not exceed 2^{Rk}.
2 The number of $x_C \in X_C$ for a given crown C with one root consisting of k branches does not exceed n^{kR+1}.
3 $|\gamma(x_C)| \leqslant \alpha^k$ where the crown C has k branches.
4 Any crown with the root (i,t) and an end $(j,0)$ contains no less than t branches.
5 The only factor in (5.8) which depends on the initial measure is the last one.

Now let us write $P(x_i^t = a)$ as a series

$$G_0 + G_1 + G_2 + \ \ldots \tag{5.9}$$

where G_k is the sum of inputs from all crowns having k branches. It follows from notes 1, 2, 3 that

$$|G_k| \leqslant \text{const} \cdot \varepsilon^k$$

where $\varepsilon = (2n)^R \alpha$.

It follows from the note 4 that the values of G_s for $s < t$ depend neither on the initial measure nor on t. This allows us to take the limit in $t \to \infty$ of every member in (5.9).

In the same way the limit

$$\lim_{t\to\infty} \tilde{\mu}(x_{i_1}^t = a_1, \ldots, x_{i_k}^t = a_k)$$

can be proved to exist, and not to depend on the initial state (as it can be represented as the sum over all crowns with k roots (i_1,t), ... (i_k,t)).

The necessary values of α do not depend on k and the convergence is uniform in the initial state. The theorem is proved.

Let us show two applications of obtained expansions: a service system and our Example 1.2.

In all the integer points of a line there are automata of which every one has two states: 0 (free) or 1 (busy). If the automaton i is busy at time t, then the next moment of the integer time it either fulfils the service with probability $(1 - \varepsilon)$, or continues it, or transmits it to its right neighbour if it is possible (i.e. right neighbour is free). Also, a flow of requests comes to the system which results in free automata getting busy with probability p. Thus we have a system with three neighbours: $V = \mathbb{Z}$, $U(h) =$

$\{h-1, h, h+1\}$, with the transition probabilities shown in Table 5.1, where asterisks stand for unimportant values. We assume that ε is small. This suggests that we take $\eta(1) = p$, $\eta(0) = 1 - p$. This results in the values of β as in Table 5.2. Of course, $\beta(0|\ldots) = -\beta(1|\ldots)$.

Now, applying the general Theorem 5.3 to this model gives.

Proposition 5.4 The cluster expansion converges provided ε is small (quick service) or $(1 - p)$ is small (frequent requests).

The other application is to our Example 1.2 (Stavskaya's problem), where transition probabilities are as in Table 5.3. We put

$$\eta(1) = \theta, \ \eta(0) = 1 - \theta.$$

Table 5.1

x^t_{i-1}	x^t_i	x^t_{i+1}	$\theta(1\|\ldots)$
0	0	*	p
*	1	0	p
*	1	1	$\varepsilon + p - \varepsilon p$
1	0	*	$\varepsilon + p - \varepsilon p$

Table 5.2

x^t_{i-1}	x^t_i	x^t_{i+1}	$\beta(1\|\ldots)$
0	0	*	0
*	1	0	0
*	1	1	$\varepsilon(1 - p)$
1	0	*	$\varepsilon(1 - p)$

Table 5.3

x^t_{i-1}	x^t_i	$\theta(1\|\ldots)$
*	0	θ
0	*	θ
1	1	1

Table 5.4

x^t_{i-1}	x^t_i	$\beta(1 \mid \ldots)$
*	0	0
0	*	0
1	1	$1 - \theta$

Then the values of β are as in Table 5.4. Of course,

$$\beta(0 \mid \ldots) = -\beta(1 \mid \ldots).$$

One can see from the definition of $\beta(.)$ that only those x_C give non-zero contributions to $\tilde{\mu}(x^t_i = 1)$ which consist of ones only, without zeros. Thus non-zero summands in $\tilde{\mu}(x^t_i = 1)$ correspond to crowns. In each of them every branch gives a factor $(1 - \theta)$ and every end gives a factor θ.

If θ is large, that is $(1 - \theta)$ is small, then β is small according to its definition and Theorem 5.3 asserts ergodicity which is already known.

Now let θ be small.

In this case the transition probabilities strongly depend on their arguments, which precludes straightforward application of the cluster method. But in this special case there is a special method. Let $\tilde{\mu}_0$ stand for the evolution measure of the Stavskaya system with the initial measure 'all zeros'.

Lemma 5.5 $\tilde{\mu}_0(x^t_i = 1)$ equals the sum of $\gamma(\mathbf{1}_C)$ over all crowns with the root (i,t) in the band $(0;t]$ filled with ones.

Proof of this lemma just repeats the proof of Lemma 5.2.

Theorem 5.6 [61]. The limit exists

$$\lim_{t \to \infty} \tilde{\mu}_0(x^t_i = 1)$$

and equals the sum of $\gamma(\mathbf{1}_C)$ over all crowns with the root (i,t). This limit is analytical in some interval $(0;\theta_0)$ and can be analytically extended to the complex domain

$$\{\theta\colon |\theta| < \theta_0, |1-\theta| < 1\}.$$

Proof

1 If a crown has m branches and n ends, then $n \leqslant m + 1$.
2 The number of crowns having one root (i,t) and n ends do not exceed

some ℓ^n (this follows from estimation of the number of boundaries of crowns).

3 A crown C with n ends and m branches has

$$\gamma(\mathbf{1}_C) = \theta^n(1-\theta)^m \leqslant \theta^n.$$

4 Any crown with the root (i,t) having an end $(j,1)$ has no less than t ends.

Now we represent

$$\tilde{\mu}(x_i^t = 1) = G_1 + G_2 + \ldots$$

where G_n is the sum of $\gamma(.)$ from all crowns with n ends. Note 4 proves that the values of G_s with $s < t$ do not depend on t. For a while let us write

$$G_n = \sum_{m=n-1}^{\infty} a_{mn}\theta^n\beta^m$$

where β is inserted instead of $(1-\theta)$, and allow β be independent of θ. In fact this sum is finite because

$$\sum_m a_{mn} \leqslant \ell^n.$$

Hence the double series ΣG_n converges in the polycircle

$$\{(\theta,\beta)\colon |\theta| \leqslant \theta_0,\ |\beta| \leqslant 1\}$$

for any $\theta_0 < \dfrac{1}{\ell}$ uniformly in t [69] and we may substitute everywhere its member by its limit in $t \to \infty$. The limit series is an analytical function of θ and β in the polycircle

$$\left((\theta,\beta)\colon |\theta| < \frac{1}{\ell},\ |\beta| < 1\right)$$

which is continuous in the closed polycircle

$$\{(\theta,\beta)\colon |\theta| \leqslant \theta_0,\ |\beta| \leqslant 1\}$$

for any $\theta_0 < \dfrac{1}{\ell}$. The substitution $\beta = 1 - \theta$ turns it into an analytical function of one argument θ in the domain

$$\left(\theta\colon |\theta| < \frac{1}{\ell},\ |1-\theta| < 1\right)$$

which is continuous in the closed domain

$$\{\theta\colon |\theta| \leqslant \theta_0,\ |1-\theta| \leqslant 1\}.$$

Just from its definition this function equals

$$\lim_{t\to\infty} \tilde{\mu}(x_i^t = 1)$$

for

$$0 \leqslant \theta < \theta_1 = \frac{1}{\ell}.$$

With $\theta \to 0$ it tends to 0 (from its continuity) which shows that our $\tilde{\mu}_0$ is not concentrated in 'all ones' and the operator is not ergodic.

Now let θ take any value from 0 to 1. In this general case, in spite of the absence of estimations providing analyticity, we can still prove convergence of the sum

$$\sum_n \sum_{m=n-1}^{\infty} a_{mn}\theta^n(1-\theta)^m,$$

existence of the limit

$$\lim_{t\to\infty} \tilde{\mu}_0(x_i^t = 1)$$

and equality of this sum and this limit. The point is that members of the sum are non-negative and for any finite t the value of

$$\tilde{\mu}_0(x_i^t = 1)$$

equals some partial sum into which every member of the infinite sum gets sooner or later.

As in Theorem 5.3, all the assertions of Theorem 5.6 can be extended to probabilities

$$\tilde{\mu}_0(x_{i_1}^t=1,\ldots,x_{i_k}^t=1).$$

Each of them has a limit in $t \to \infty$ which is representable as a cluster expansion the sum of which is analytical in some interval $(0;\theta_0)$ and can be analytically extended to the complex domain mentioned in the statement of Theorem 5.6.

The initial condition 'all zeros' is not quite necessary in Theorem 5.6. A homogeneous independent measure μ with $\mu(x_i = 1) \leqslant \theta$ will do too. See also [98] where there are some stronger results.

Some other techniques including correlation equations for evolution measure prove analyticity of the $\tilde{\mu}_0$ invariant measure and existence of its cluster expansion in some neighbourhood of $\theta = 0$ [100].

The method of this chapter can be generalised in various ways. First, it works with obvious alterations for the case when transitional probabilities depend on several previous times $t-1, t-2, \ldots, t-T$ where T is fixed.

Second, this method can be generalised to some non-independent operators too, namely the so-called Markov chains of automata [15]. As before, we speak of $\tilde{\mu}$ where x_i^t is the state of the i automaton at t time. We order points $(i,t) \in \mathbb{Z}^2$ by the rule: $(j,s) < (i,t)$ if $s < t$ or $s = t$ and $j < i$ and let any state x_i^t depend in a certain random way on some of those x_j^s for which $(j,s) < (i,t)$. But this dependence is local in the sense that x_i^t depends only on a finite number of its predecessors, namely those for which

$$\max(|s-t|, |i-j|) \leqslant \text{const.}$$

Another generalisation of cluster expansions is to use

$$\theta(x_i^t|x^{t-1}) = p_0(x_i^t|x_i^{t-1}) + \beta(x_i^t, x^{t-1})$$

instead of (5.1), where $p_0(.|.)$ is the matrix of some irreducible non-periodic Markov chain. If $|\beta| < \alpha$ and α is small, this means weak interdependence between automata. In this case a branch is defined as before, but some definitions have to take new forms:

Definition 5.1′

(1) Branch B_{ju} is said to be *following* the point (j,t) if $u \leqslant t$. Branch B_{ju} is said to be *following* the branch B_{it} if it follows some of its ends, that is, if $u < t$ and $|j - i| \leqslant R$.
(2) A *crown* with the root (i,t) is a finite set of branches one of which follows (i,t) and each of the others follows some branch of the crown. Ends of a crown are such ends of its branches for which all points (j,v), where $v < u$, do not belong to the crown.
(3) A *crown with several roots* is defined analogously: it consists of a finite number of branches iteratively following the points

$$(i_1,t), \ldots, (i_k,t)$$

As before, $\gamma(x_C)$ is the product of factors which include $\beta(.,.)$ corresponding to the branches, and $\eta(.)$ factors corresponding to the ends of the crown C where η is now the stationary distribution of the Markov chain with transition probabilities $p_0(x|y)$. But now there is a new factor $p_0^{(u-v)}(x_j^u|x_j^v)$ for every pair of points (j,v), (j,u) such that (j,u) and (j,v) belong to the crown and $(j,v+1)$, $\ldots$ $(j,u-1)$ do not belong to the crown. Here $p_0^{(n)}(x|y)$ is the probability of transition from y to x in n steps.

Lemma 5.2 literally fits here and a theorem like Theorem 5.3 is provable.

Chapter 6
Random walk operators: ergodicity

Although our main concern is with independent operators, we use some other operators as tools. Here we introduce and begin to examine non-independent 'random walk' operators introduced by A. L. Toom. Throughout this chapter we assume $X = \{0,1\}^{\mathbb{Z}}$ and $0 \prec 1$ and speak of monotonicity in the sense of Chapter 2.

Proposition 6.1 Let W be a monotone operator from X to X and $P \in \mathscr{P}_1$ (which means that P is from X to X and conserves the measure δ_1 concentrated in the state 'all ones'). If $W \prec P^\tau$ with some τ and $\forall\, x \in X$:

$$\lim_{t \to \infty} xPW^t = \delta_1$$

then P is ergodic.

Proof is easy.

We are going to use this proposition to prove ergodicity of some independent operators P.

But first we have to define our tool: hence W stands for a random walk operator. Let us term a segment of $\mathbb{Z}$

$$\{k, k+1, \ldots, k+s-1\} \subset \mathbb{Z}$$

a 'massif' of ones of the length S for a given $x \in X$ if $x_{k-1} = x_{k+s} = 0$, $x_k = x_{k+1} = \ldots = x_{k+s-1} = 1$. Infinite massifs of ones are defined in an analogous way.

A random walk operator depends on some parameters; thus before its definition we must choose two natural numbers r and s such that $r \geqslant 2s$. We also need two probability distributions F and G on the set

$$\{-s, -s+1, \ldots, s-1, s\} \subset \mathbb{Z}$$

Definition 6.2 [83]. $W = W_1 W_2$ where W_1 and W_2 are operators from X to X.

First W_1 is performed; W_1 is deterministic and depends on the parameter r. In fact it eliminates all massifs of ones which are shorter than r. In other words $(xW_1)_h = 1$ if and only if there is such i that $h \in [i, i+r-1]$ and $x_i = \ldots = x_{i+r-1} = 1$.

Now W_2 works. Informally, it moves all the left-hand ends of the massifs of ones according to the distribution F and moves all the right-hand ends of the massifs of ones according to G, and all these movements are mutually independent.

Formally, let $\xi_1, \xi_2, \ldots$ and $\eta_1, \eta_2, \ldots$ be two countable families of independent random variables distributed according to F and G respectively. Let all massifs of ones in x be somehow enumerated and let $[i_k, j_k]$ be the kth massif. W_2 transforms x into the random $y \in X$ by the rule: $y_h = 1$ if and only if

$$\exists k\colon i_k + \xi_k \leqslant h \leqslant j_k + \eta_k.$$

Proposition 6.3 W is monotone.

Proof. Monotonicity of W_1 is evident. It remains to prove that W_2 is monotone on the subset XW_1 of X:

$$(x, y \in XW_1 \ \&\ x \prec y) \Rightarrow xW_2 \prec yW_2.$$

This can be proved by the coupling technique. Let y have one massif of ones $[i,j]$. Let $[i_1, j_1], \ldots, [i_k, j_k]$ be those massifs of ones of x, which belong to $[i,j]$. We may and shall identify that ξ variable which moves i in the application of W_2 to y with that ξ variable which moves i_1 in the application of W_2 to x. Similarly we identify those η variables which move j and j_k in the application of W_2 to y and x. This being done, the realisation of xW_2 is always no more than the realisation of yW_2, whence

$$xW_2 \prec yW_2$$

Note that the condition $x, y \in W_1$ was essential here. Indeed, W_2 is not monotone when applied to the whole of X.

For y having many massifs of ones, the proof is analogous.

The values of $\mathbb{E}\xi$ and $\mathbb{E}\eta$ are the most important determinants of W's behaviour. Indeed, in a long time t the left-hand and right-hand ends of long massifs of ones move at about $t \cdot \mathbb{E}\xi$ and $t \cdot \mathbb{E}\eta$. With this kept in mind, we term W 'extensive' if $\mathbb{E}\xi < \mathbb{E}\eta$.

The next proposition is a simple corollary of the classical random walk techniques. It treats results of iterative application of an extensive W to a state having one massif of ones. All the resulting measures at times $t = 1, 2, \ldots$ are concentrated in states having no more than one massif $[i_t, j_t]$.

[If $x^t = \mathbf{0}$ then i_t and j_t are undefined). Of course, i_t and j_t perform random walks on $\mathbb{Z}$ with independent single steps distributed as ξ and η.

Proposition 6.4 Let us have an extensive W and denote

$$c_1 = (3\cdot\mathbb{E}\xi + \mathbb{E}\eta)/4,\ c_2 = (\mathbb{E}\xi + 3\cdot\mathbb{E}\eta)/4.$$

Let $\tilde{\mu}$ be the evolution measure produced by W with the initial state x^0

$$x_h^0 = \begin{cases} 1 & \text{if} \quad i_0 \leqslant h \leqslant j_0 \\ 0 & \text{otherwise} \end{cases}$$

where $j_0 - i_0 > r$.

There are such u, v (which depend on W, but not on i_0, j_0) that for any m, n such that $r + m + n \leqslant j_0 - i_0$, x^t will never become $\mathbf{0}$ and the following inequalities

$$\begin{aligned} i_t &\leqslant i_0 + m + c_1 t = i_t^* \\ j_t &\geqslant j_0 - n + c_2 t = j_t^* \end{aligned} \tag{6.1}$$

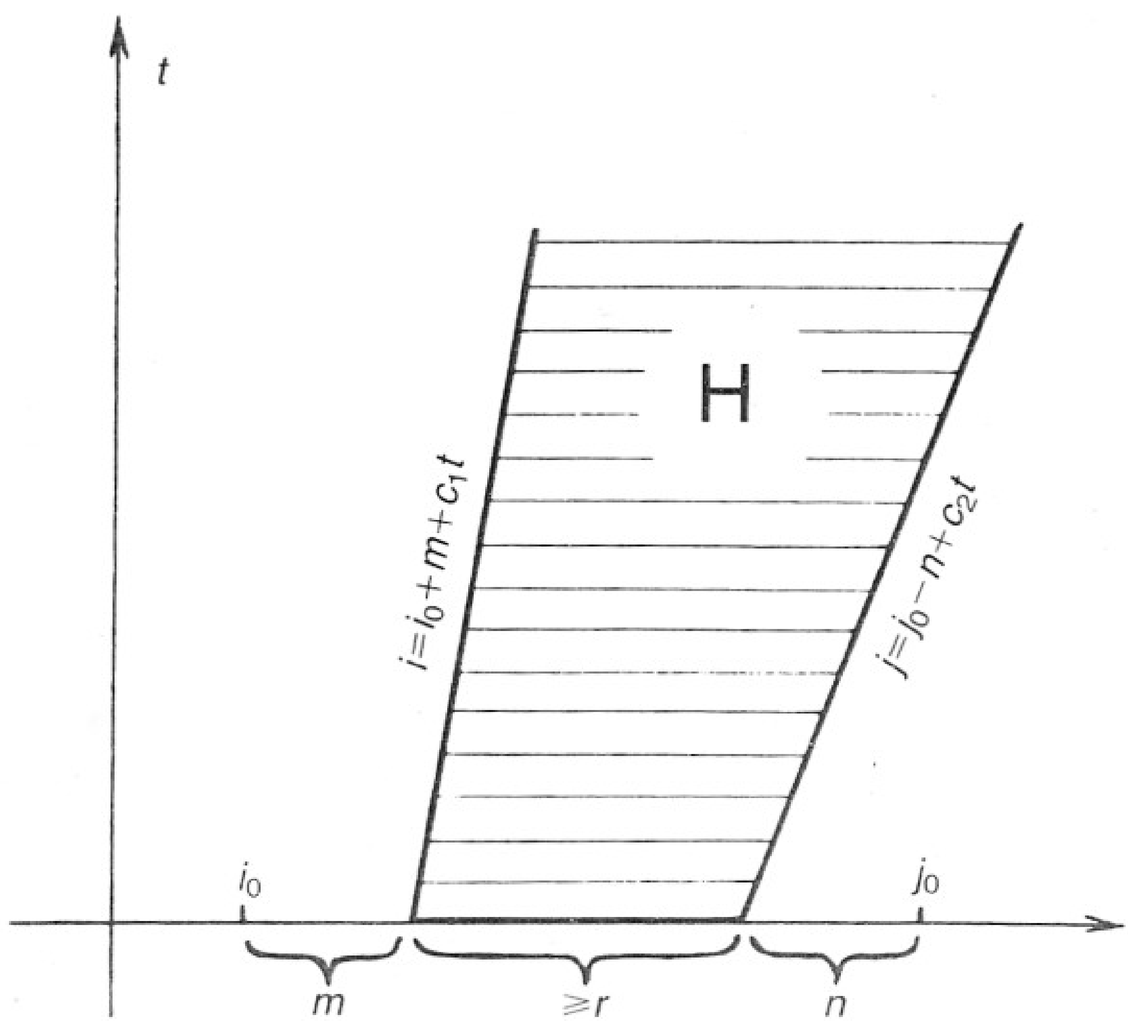

Fig. 6.1 Domain H. See Proposition 6.4.

hold for all $t \geqslant 0$ with probability no less than $1 - u^m - v^n$.

Informally this means that in the case $\mathbb{E}\xi < \mathbb{E}\eta$ long initial massifs of ones will never die and will even grow with large probabilities, thus filling the domain H of the evolution space, which is shown in Figure 6.1.

Proof directly follows from [19, ch. 14, 8].

Lemma 6.5 (A. L. Toom). Let W be extensive. Let the initial measure μ be such that for any ℓ the probability to find in a segment $[i,j] \subset \mathbb{Z}$ some segment of the length ℓ filled with ones tends to 1 with $(j - i) \to \infty$ uniformly in i and j. Then

$$\lim_{t\to\infty} \mu W^t = \delta_{\mathbf{1}}.$$

Proof. It is sufficient to prove that $\tilde{\mu}(x_h^t = 1)$ tends to 1 with $t \to \infty$ uniformly in $h \in \mathbb{Z}$.

Let us fix $\varepsilon > 0$ and find such t_0 that

$$t \geqslant t_0 \Rightarrow \tilde{\mu}(x_h^t = 1) \geqslant 1 - \varepsilon.$$

Proposition 6.4 provides

$$\tilde{\mu}(x_h^t = 1 | x_{[i_0,j_0]} = 1) \geqslant 1 - u^m - v^n$$

for $(0,t) \in H$ where H is the infinite domain shown in Figure 6.1. Choose m, n such that

$$u^m + v^n \leqslant \varepsilon/2.$$

Now take d such that the probability of finding in any segment $[i,i+d-1]$ some segment of length $\ell = r + m + n$ filled with ones exceeds $1 - \frac{\varepsilon}{2}$.

Now denote

$$H_{[i_0,j_0]} = \{(k,t)\colon i_t^* \leqslant k \leqslant j_t^*\} \tag{6.2}$$

the domain in the evolution space $\mathbb{Z} \times \mathbb{Z}_+$ filled with ones with large probability according to Proposition 6.4, where i_t^* and j_t^* are defined in (6.1). The typical form of $H_{[i_0,j_0]}$ is shown in Figure 6.1 as H. Simple geometrical consideration shows that for any fixed ℓ, i, d all $H_{[i_0,j_0]}$ with $[i_0,i_0+\ell-1] \subset [i,i+d-1]$ have a non-empty common part. Taking (h,t) in this part proves the lemma.

Let us term an operator $P \in \mathscr{P}_1$ on $X = \{0;1\}^{\mathbb{Z}}$ extensive if there is such τ and such an extensive random walk operator W that $W \prec P^\tau$. According to Lemma 6.5 and Proposition 6.1 any extensive P is ergodic. The reverse is partly true too:

Proposition 6.6 (A. M. Leontovitch). Let P be such an operator on $X = \{0;1\}^{\mathbf{Z}}$ that for any $\alpha > 0$ there is such t that

$$\mu P^t(x_i = 0) < \alpha/t$$

uniformly in $\mu \in \mathcal{M}(x)$ and $i \in \mathbb{Z}$. Then P is extensive. Proof is in [55].

Note 6.7 Assuming the conditions of Lemma 6.5 and the measure $\mu \neq \delta_0$ being homogeneous and independent, μW^t tends to δ_1 exponentially. Thus, in this case only two types of convergence are possible:

if P is extensive, $\delta_0 P^t(x_i = 1)$ tends to 1 exponentially;
if P is not, there is such $\alpha > 0$ that

$$\forall t\colon\ \delta_0 P^t(x_i = 0) \geqslant \alpha/t.$$

Thus, if we want to prove ergodicity of a given $P \in \mathcal{P}_1$ on $\{0;1\}^{\mathbb{Z}}$ we should seek τ and W such that $W \prec P^\tau$. This can be done by applying P^τ to a state having one infinite massif of ones and reducing all the resulting states to those having only one infinite massif of ones too by turning into zeros those ones which do not belong to it. Doing this for operators $P \in \mathcal{P}_1$ on the graph Γ_1 in the case $\tau = 1$ provides us with such a W, which has

$$\mathbb{E}\xi = 1 - \frac{\theta_{01}}{1 - \theta_{00}},\ \mathbb{E}\eta = \frac{\theta_{10}}{1 - \theta_{00}}.$$

So, this W is extensive if

$$\theta_{00} + \theta_{01} + \theta_{10} > 1.$$

This inequality is sufficient for P to be ergodic. In particular, the operator of Example 1.2 thus gets proved ergodic in the range $\varepsilon > \frac{1}{3}$. Taking larger τ gives better estimations. In the special case of Example 1.2: taking $\tau = 2$ proves ergodicity for $\varepsilon > 0.328$ and taking $\tau = 3$ proves ergodicity for $\varepsilon > 0.324$. Thus we have for the critical value in Example 1.2:

$$\varepsilon^* \leqslant 0.324.$$

In percolation theory some constructions have been used based on ideas similar to our random walk operators [17]. In particular, J. Bishir [91] proved the estimation $\varepsilon^* \leqslant \frac{1}{3}$ for our Example 1.2.

Quite another way to prove uniqueness of invariant measure was advanced by A. M. Leontovitch. It has not yet been developed enough to be treated in a separate chapter, and we thus consider here what is known about it. Its key idea is to use the rule of contraries.

A homogeneous normed measure μ on $X = \{0;1\}^{\mathbb{Z}}$ can be defined by its values:

$$\mu_{z_1 \ldots z_k} = \mu\{x\colon x_{i+1}=z_1,\ldots,x_{i+k}=z_k\}. \tag{6.3}$$

Moreover, even these values are not independent and just those of the (6.3) values in which $z_1 = z_k = 0$ define μ. These are some of them:

$$\mu_0, \mu_{00}, \mu_{000}, \mu_{010}, \ldots \tag{6.4}$$

For example,

$$\begin{aligned} \mu_1 &= 1 - \mu_0, \\ \mu_{01} &= \mu_{10} = \mu_0 - \mu_{00}, \\ \mu_{11} &= \mu_1 - \mu_{01}, \end{aligned}$$

and so on.

Finally, μ of the single state 'all ones' is

$$1 - (\mu_0 + \mu_{010} + 2\mu_{0110} + 3\mu_{01110} + \ldots).$$

Now let P be a homogeneous operator on X. Let $P \in \mathscr{P}_1$, in which case the measure δ_1 concentrated in 'all ones' is invariant. Suppose P has another measure μ in which $\mu_0 > 0$. The equation $\mu = \mu P$ can be rewritten as a homogeneous system of equations for the values (6.4).

This system has a non-trivial non-negative solution, for which the series

$$\mu_0 + \mu_{010} + 2\mu_{0110} + 3\mu_{01110} + \ldots \tag{6.5}$$

converges, if and only if P has another homogeneous invariant measure μ as well as δ_1. For P monotone, the uniqueness of δ_1 as invariant measure implies ergodicity.

Let us apply these considerations to our Example 1.2. These are the first few equations of the infinite system equivalent to $\mu = \mu P$:

$$\mu_0 = (1 - \varepsilon)(2\mu_0 - \mu_{00}), \tag{6.6}$$

$$\mu_{00} = (1 - \varepsilon)^2(\mu_0 + \mu_{010}), \tag{6.7}$$

$$\mu_{000} = (1 - \varepsilon)^3(\mu_{00} + 2\mu_{010}), \tag{6.8}$$

$$\mu_{010} = \varepsilon(1 - \varepsilon)^2(\mu_{00} + 2\mu_{010}) + (1 - \varepsilon)^2\mu_{0110}, \tag{6.9}$$

$$\mu_{0000} = (1 - \varepsilon)^4(\mu_{000} + \mu_{010} + \mu_{0010} + \mu_{01010}), \tag{6.10}$$

$$\mu_{0010} = (1 - \varepsilon)^3\mu_{0110} + \varepsilon(1 - \varepsilon)^3(\ldots), \tag{6.11}$$

$$\mu_{0110} = 2\varepsilon(1 - \varepsilon)^2\mu_{0110} + (1 - \varepsilon)^2\mu_{01110} + \varepsilon^2(1 - \varepsilon)^2(\ldots) \tag{6.12}$$

and so on.

Making inferences from these equations results in reducing the range of ε. Just the first equation gives

$$\mu_{00} = \frac{1 - 2\varepsilon}{1 - \varepsilon}\mu_0.$$

But both μ_0 and μ_{00} should be non-negative, which is impossible if $\varepsilon > \frac{1}{2}$. This proves $\varepsilon^* \leqslant \frac{1}{2}$. Also μ_{00}, μ_{010}, μ_{0110} can be expressed in terms of μ_0 using (6.6)–(6.9), which results in

$$\varepsilon^* \leqslant 0.42, \ \varepsilon^* \leqslant 0.38, \ \varepsilon^* \leqslant 0.35$$

respectively. We cannot express other values of (6.4) in terms of μ_0. However, better estimations for ε^* can be obtained. For example, (6.12) yields

$$(1 - \varepsilon)^2\mu_{01110} = (1 - 2\varepsilon(1 - \varepsilon)^2)\mu_{0110} - \\ - \varepsilon^2(1 - \varepsilon)^2(\mu_{000} + \mu_{010} + \mu_{0010} + \mu_{0100} + \mu_{01010}).$$

Substituting here our formulae for μ_{0110}, μ_{000} and μ_{010} and expressing μ_{0010} and μ_{0100} in terms of μ_{000}, μ_{010}, μ_{0110}, μ_{01010} with (6.11) we obtain $\mu_{01110} < 0$ for $\varepsilon > 0.33$.

Thus $\varepsilon^* \leqslant 0.33$. We see that this method gives about the same numerical estimations for Example 1.2 as the random walk operators method.

Now let us discuss all the estimations of Example 1.2 operators we have got. Random walk operators provide upper estimations of ε^* which as $\tau \to \infty$ tend to some $\varepsilon^*_{\text{walk}}$ which is the infimum of those ε with which the operator in question is extensive. In the range $\varepsilon > \varepsilon^*_{\text{walk}}$ our operator is ergodic and μP^t tends to 'all ones' exponentially. Clearly $\varepsilon^* \leqslant \varepsilon^*_{\text{walk}}$. If they are not equal, our operator in the range $\varepsilon^* < \varepsilon < \varepsilon^*_{\text{walk}}$ has only one invariant measure $\delta_{\mathbf{1}}$ but $\mathbf{0}P^t$ tends to $\delta_{\mathbf{1}}$ slowly: $\delta_{\mathbf{0}}P^t(x_i = 0)$ is not $o(1/t)$.

We can also define $\varepsilon^*_{\text{syst}}$ which is the infimum of those $\varepsilon_{(6.6)\ldots}$ with which the infinite system has no non-trivial non-negative solution. Clearly $\varepsilon^* \leqslant \varepsilon^*_{\text{syst}}$. If they are not equal, in the range $\varepsilon^* < \varepsilon < \varepsilon^*_{\text{syst}}$ the system (6.6)... has a solution, but the series (6.5) diverges. However, it seems most plausible that

$$\varepsilon^* = \varepsilon^*_{\text{walk}} = \varepsilon^*_{\text{syst}}.$$

Chapter 7
Ergodicity at the simplest graph

This chapter applies all the methods described above to the particular case of homogeneous independent operators on the graph Γ_1 with $X = \{0;1\}^{\mathbb{Z}}$. This family of operators has four parameters

$$\theta(x_h^{t+1} = 1 \mid x_{h-1}^t, x_h^t)$$

denoted here:

$$\theta_{00},\ \theta_{01},\ \theta_{10},\ \theta_{11}$$

where the two indices stand for x_{h-1}^t, x_h^t. For simplicity we reduce these four to three by assuming $\theta_{01} = \theta_{10}$.

First we apply the results of Chapter 4. The characteristic polynomial of our operator P for the standard functions χ_K is

$$P\chi_i = \theta_{00} + (\theta_{10} - \theta_{00})(\chi_{i-1} + \chi_i) + (\theta_{11} + \theta_{00} - 2\theta_{10})\chi_{i-1}\chi_i$$

and the condition (4.5) takes the form

$$\theta_{00} + 2|\theta_{10} - \theta_{00}| + |\theta_{11} + \theta_{00} - 2\theta_{10}| < 1. \qquad (7.1)$$

In general an inequality

$$f_0(v) + \sum_{n=1}^{N} |f_n(v)| < C$$

where $v \in \mathbb{R}^d$ and all $f_0, f_1, \ldots, f_n$ are linear, is equivalent to the system of 2^N inequalities

$$f_0(v) + \sum_{n=1}^{N} \delta_n f_n(v) < C$$

where every $\delta_n = \pm 1$. In our case (7.1) is equivalent to the system of four inequalities

$$\begin{aligned} &\theta_{11} < 1, \\ &2\theta_{00} + \theta_{11} > 4\theta_{10} - 1, \\ &4\theta_{00} + \theta_{11} < 4\theta_{10} + 1, \\ &2\theta_{00} - \theta_{11} < 1. \end{aligned}$$

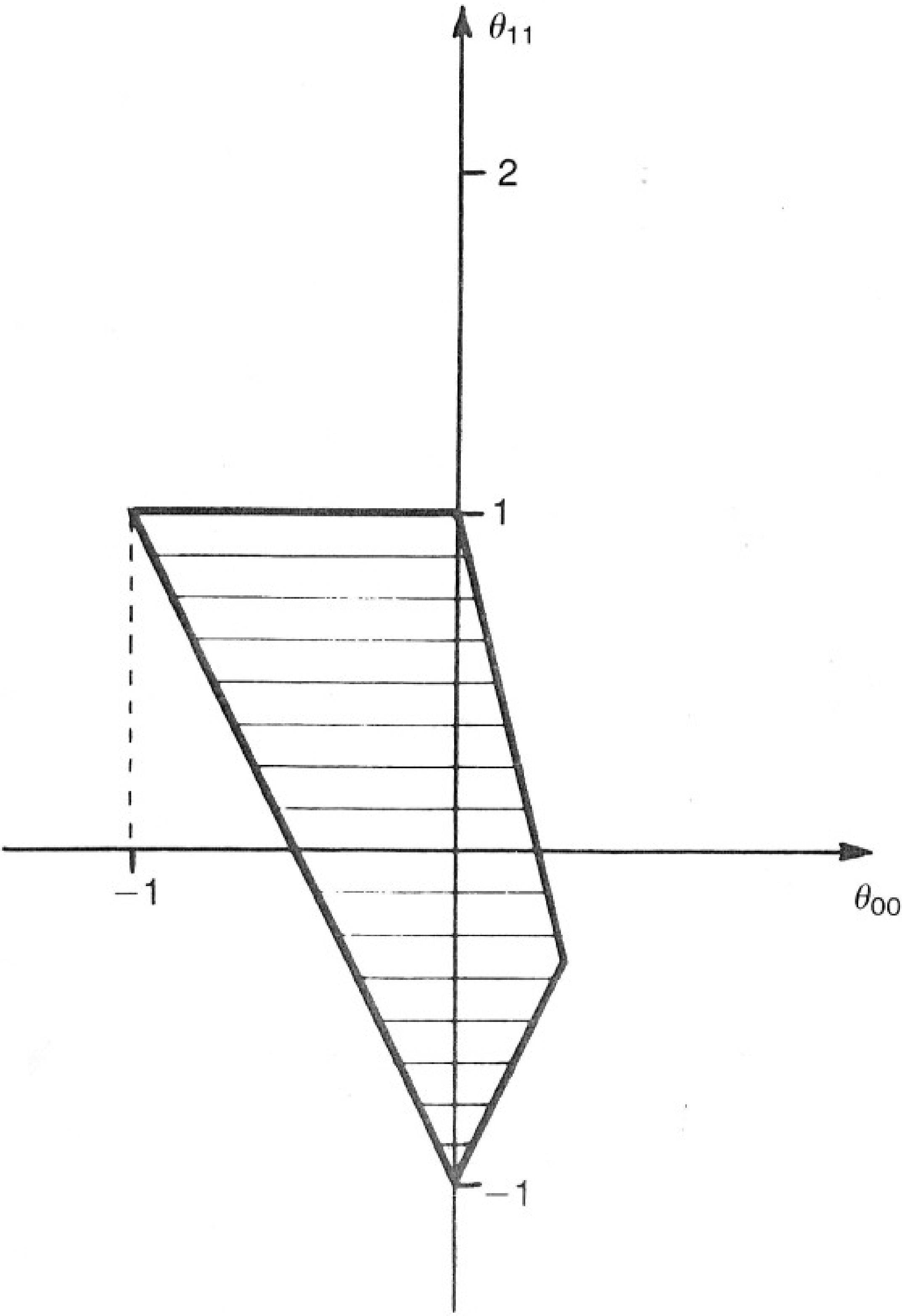

Fig. 7.1 Polygon S_0.

This is some polyhedron in the cube. Its section by the plane $\theta_{10} = 0$ is the quadrilateral S_0 shown in Figure 7.1. For any θ_{10} fixed, the corresponding section is the intersection of the square and the image of S_0 under the homothety with centre (1,1) and coefficient $1 - \theta_{10}$.

But the functions χ_K chosen are no better than any arbitrary χ'_K. Let us take χ'_K with $a = -1$ and $b = 0$ which is the opposite of the χ_K case. Now the characteristic polynomial is

$$P\chi'_i = 1 - \theta_{11} + (\theta_{10} - \theta_{11})(\chi'_{i-1} + \chi'_i) + (2\theta_{10} - \theta_{11} - \theta_{00})\chi'_{i-1}\chi'_i.$$

This leads us to the system

$$\begin{aligned} &\theta_{00} > 0, \\ &\theta_{00} + 2\theta_{11} < 4\theta_{10}, \\ &\theta_{00} + 4\theta_{11} > 4\theta_{10}, \\ &\theta_{00} - 2\theta_{11} < 0. \end{aligned}$$

The section of this polyhedron by the side $\theta_{10} = 1$ is S'_1 shown in Figure 7.2. Any section of it by the plane parallel to this side is the intersection of the square with the image of S'_1 under homothety with centre (0.0) and coefficient θ_{10}.

Now let us take $a = -1$, $b = 1$. In this case

$$\begin{aligned} P\chi'_i = &\tfrac{1}{2}(\theta_{00} + \theta_{11} + 2\theta_{10} - 2) + \\ &+ \tfrac{1}{2}(\theta_{11} - \theta_{00})(\chi'_{i-1} + \chi'_i) + \tfrac{1}{2}(\theta_{00} + \theta_{11} - 2\theta_{10})\chi'_{i-1}\chi'_i. \end{aligned}$$

The condition (4.7) in this case is equivalent to the system

$$\begin{aligned} &0 < \theta_{00} < 1, \\ &0 < \theta_{11} < 1, \\ &|\theta_{00} - \theta_{11}| < 1 - |2\theta_{10} - 1|. \end{aligned}$$

The section of this polyhedron T by any plane $\theta_{10} =$ const is a hexagon like the one in Figure 7.3.

We have proved ergodicity in the union $S \cup S' \cup T$. Looking at the typical sections in Figure 7.4 we see that our set has a triangular caverne R defined by the inequalities

$$\begin{aligned} &4\theta_{00} + \theta_{11} > 4\theta_{10} + 1, \\ &\theta_{11} - \theta_{00} > 2\theta_{10}, \\ &\theta_{11} < 1. \end{aligned}$$

However, using functions χ'_K with $-1 < a < 0$, $b = 1$ we can prove ergodicity for every point of R too. We can do the same for the symmetric triangle R' defined by

$$\begin{aligned} &\theta_{00} + 4\theta_{11} \leqslant 4\theta_{10}, \\ &\theta_{11} - \theta_{00} \geqslant 2 - 2\theta_{10}, \\ &\theta_{00} > 0, \end{aligned}$$

by taking $a = -1$, $0 < b < 1$.

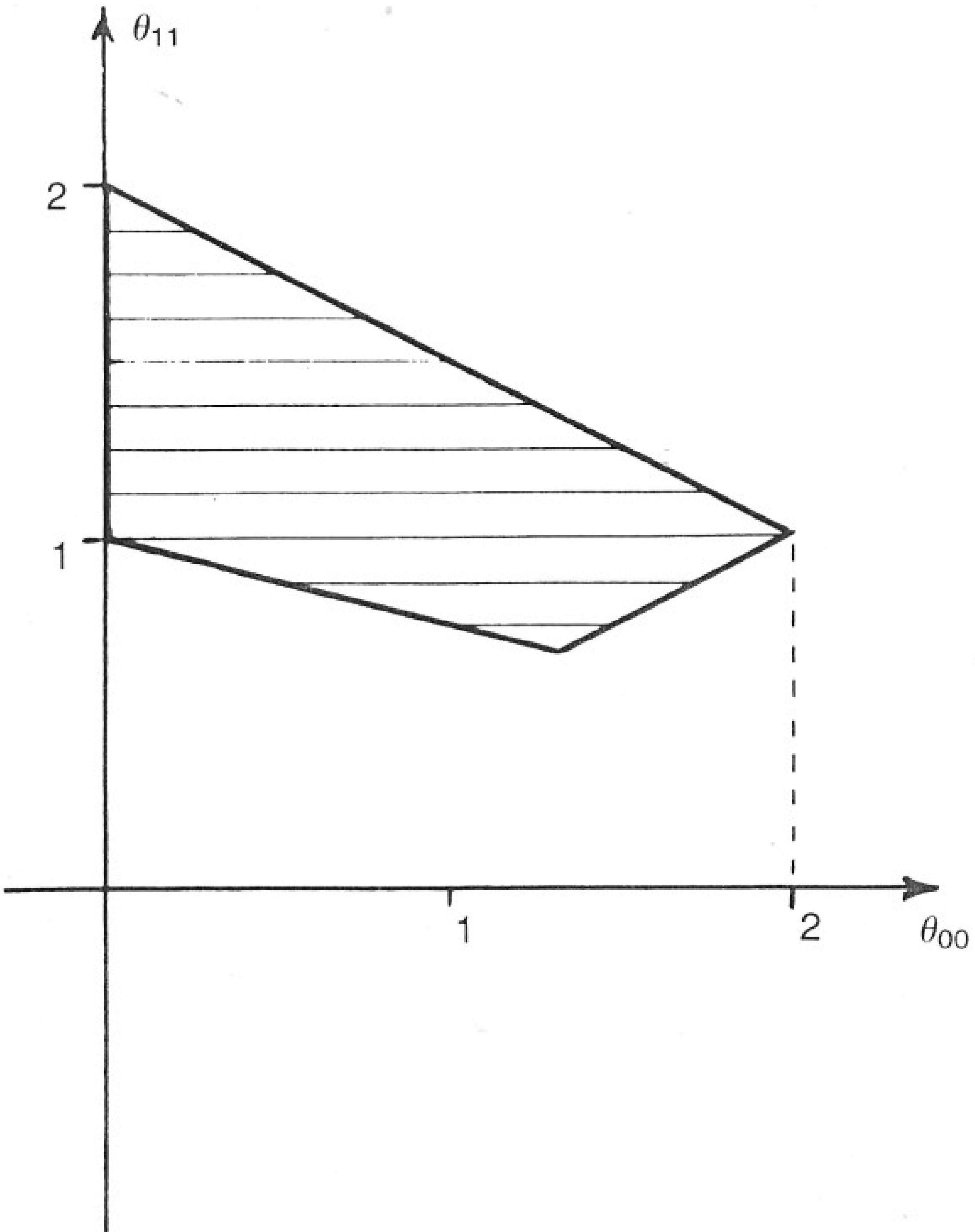

Fig. 7.2 Polygon S_1'.

Thus, the characteristic polynomials method proves ergodicity at least in the polyhedron defined by the inequalities

$$\left.\begin{array}{l} 0 < \theta_{00},\ \theta_{10},\ \theta_{11} < 1, \\ \theta_{11} > \theta_{00} - 2\theta_{10}, \\ \theta_{11} > \theta_{00} - 2(1 - \theta_{10}). \end{array}\right\} \tag{7.2}$$

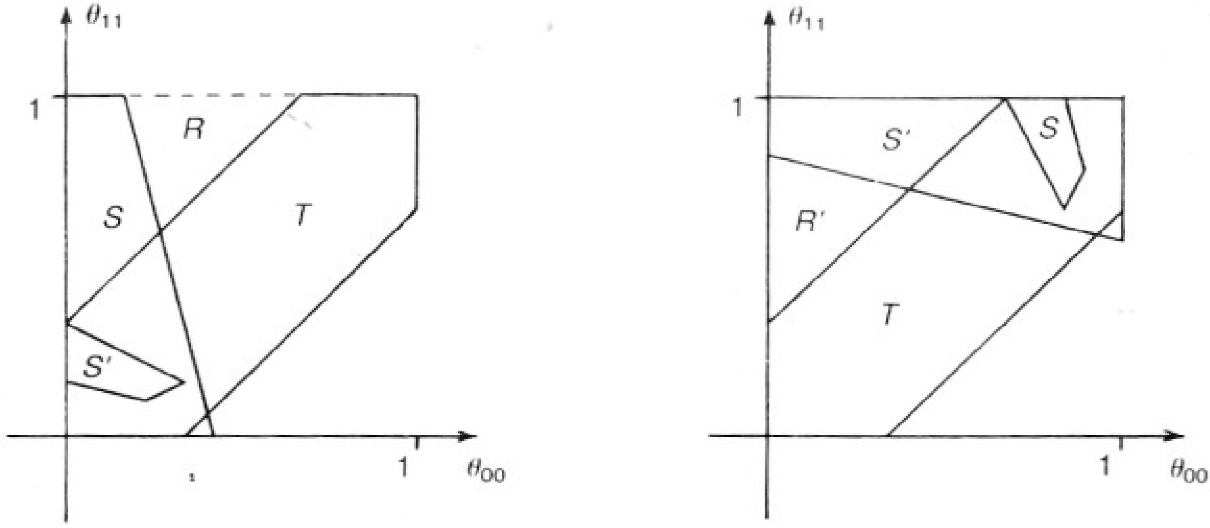

Fig. 7.4 Sections of $S \cup S' \cup T$ with planes $\theta_{10} = 1/6$ and $\theta_{10} = 5/6$.

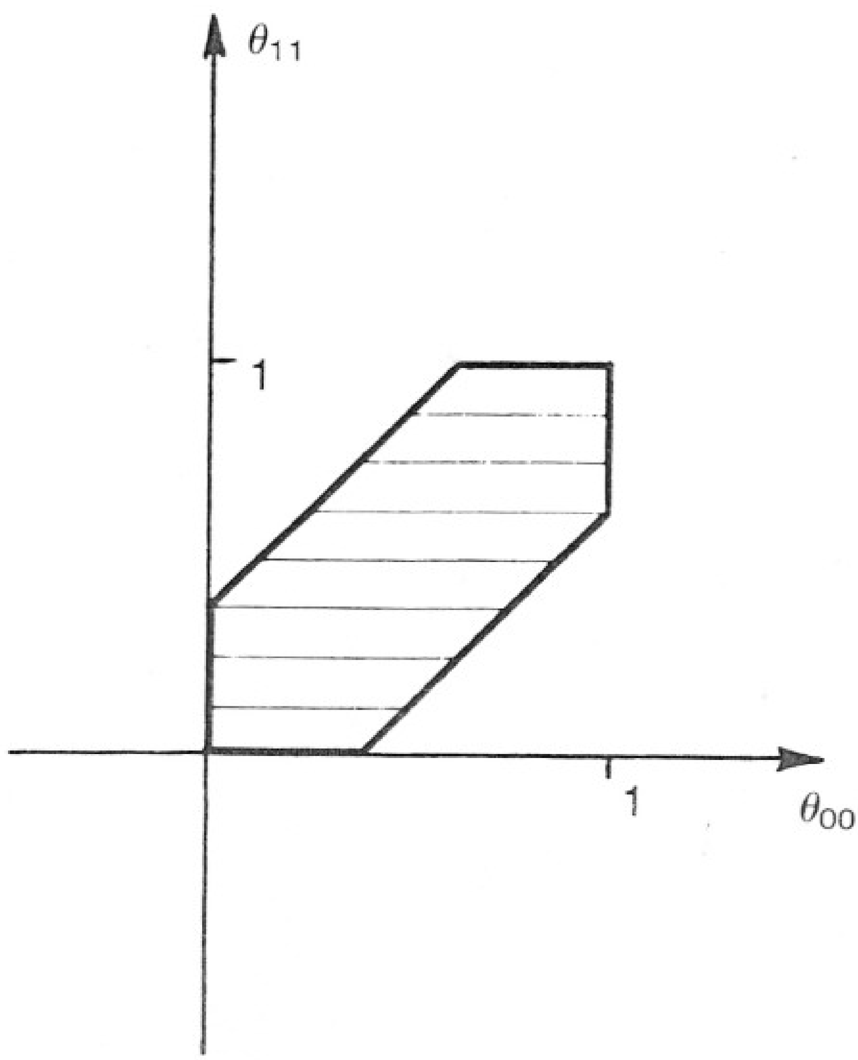

Fig. 7.3 Section of T with plane $\theta_{10} = 1/5$ (and also $\theta_{10} = 4/5$).

Now let us apply the results of Chapter 3. Corollary 3.5 and results of Chapter 6 prove ergodicity if

$$\max\{|\theta_{xy} - \theta_{zt}|\} + 2\max\{|\theta_{11}-\theta_{10}|,|\theta_{00}-\theta_{10}|\} < 2, \qquad (7.3)$$

where the left-hand maximum is over all $x, y, z, t \in \{0;1\}$ (see Fig. 7.5).

Thus we have proved ergodicity in the union of (7.2) and (7.3) which occupies more than 0.9 of the cube's volume. But the neighbourhoods of two vertices

$$\theta_{00} = 1, \ \theta_{11} = 0, \ \theta_{10} = 0 \text{ or } 1$$

do not belong to it and we have no means to prove or disprove ergodicity there.

However, let us repeat the conjecture expressed in [14]: P is ergodic in all inner points of the cube. Computer simulation supports this con-

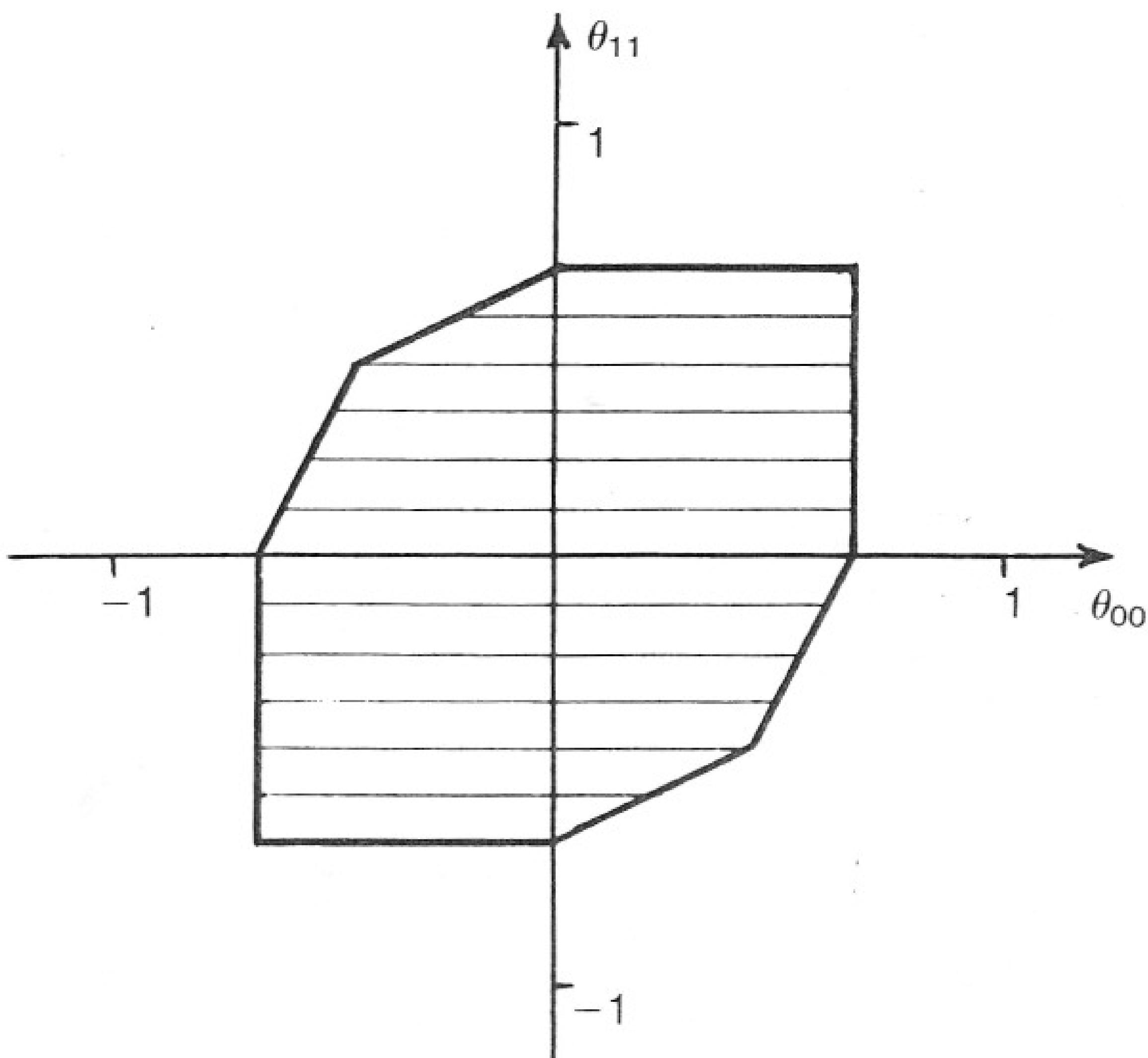

Fig. 7.5 Section of the polyhedron (7.3) with plane $\theta_{10} = 0$. For other values of θ_{10} this polygon must be shifted by vector $(\theta_{10}, \theta_{10})$.

jecture. The paper [59] describes such simulation for many inner points of the four-dimensional cube with various values of the four variables θ_{00}, θ_{01}, θ_{10}, θ_{11}. According to the results of computation, the limit $t \to \infty$ values of $\tilde{\mu}(x_i^t = 0)$ appear independent from the initial state and well approximated by the 'chaos method' computations. (The 'chaos method' was described in Chapter 2.)

Note that (7.2) implies, in particular, that any non-degenerate monotone operator on $\{0,1\}^{\mathbf{Z}}$ with symmetric dependence on two neighbours is ergodic. (In the case of continuous time the analogous result has been proved for three neighbours, see [28] and theorem 3.14 in [48, Ch. III].)