Contours, convex sets, and cellular automata

André Toom
UFPE, department of statistics, Recife, PE, Brazil
E-mail toom@bernoulli.de.ufpe.br, toom@member.ams.org

The course will be delivered in Portuguese
at the 23-th Colloquium of Brazilian Mathematicians

Content

Foreword.

The world in which we live is full of hazards and consists of many parts. It is only natural that some of its models are random and multicomponent. Mathematical systems with these properties go under various names including Interacting particle systems, Markov processes with local interaction etc. Their study started about thirty years ago and now is going on a large scale. Systems of this sort always involve some "space", where the components are placed, some variable named "time" and some set of values of each component, all of which may be continuous or discrete. On top of all that we have to choose some way of interaction between components and we can do it in various ways. So we have a lot of opportunities and the colorful phrase "programmable matter" [Tof+Mar] is very appropriate: it helps to see why interacting systems attract much attention. Whenever we define such a system, we create a world, which has some features of the real world and we can experiment with as much as we please.

Liggett's famous monograph [Liggett] contributed a lot to the study of interacting particle systems with continuous time. In our case time is discrete - as well as space and the set of states of every component. In result, the basic definitions are conceptually simple, which allows us to avoid heavy definitions and go straight to non-trivial problems and relations between different branches of mathematics. This course concentrates on the use of the contour method, convex sets and algorithms to study cellular automata.

In spite of their conceptual simplicity, cellular automata demonstrate a plethora of interesting phenomena. In a cellular automaton with discrete time and positive transition probabilities, any local event can happen with a positive probability (which is not true for systems with continuous time where only one change can occur at a time). If, in spite of this, a cellular automaton is non-ergodic (and we shall present examples of this), it is a convincing analog of phase transitions, which have many forms in the real world (freezing, melting, evaporation, condensation etc.) and are among the most important natural phenomena. We are especially interested in non-trivial (that is, strictly between zero and one) critical values of parameters, when properties of the process on the opposite sides of a critical value are qualitatively different, thus imitating natural phase transitions where the structure of matter changes qualitatively when temperature continuously changes across freezing or condensation points.

The author's main intention was to give to the students who will read this text or attend the course some taste of rigorous mathematical work in an area connected with applications. The present text consists of 5 chapters and every chapter contains at least one theorem about cellular automata. All these theorems are proved, at least for some special cases. Other statements are called lemmas or propositions. Some of them are also proved, other proofs are left to the reader. To make the text as self-sufficient as possible, we prove even such a classical statement as Helly's theorem, at least for that special case for which we need it. Besides a few classical theorems, such as undecidability of halting problem for Turing machines, most of the results included here can be found in the surveys [Discr] and [Cell], to which we refer instead of original papers, some of which are difficult to find or read. Our list of references is heterogeneous and *ad hoc*, it does not pretend for any system.

There are many solved and still more unsolved problems in this area. We selected a few of them, which are closest to our notions, and placed them at the end of every chapter as notes along with a few exercises for those students who want to get some hands-on experience of doing mathematics in this area.

1. Percolation: the first example of phase transition

The purpose of this chapter is to introduce the ideas of phase transition, critical value and contour estimation on the simplest examples.

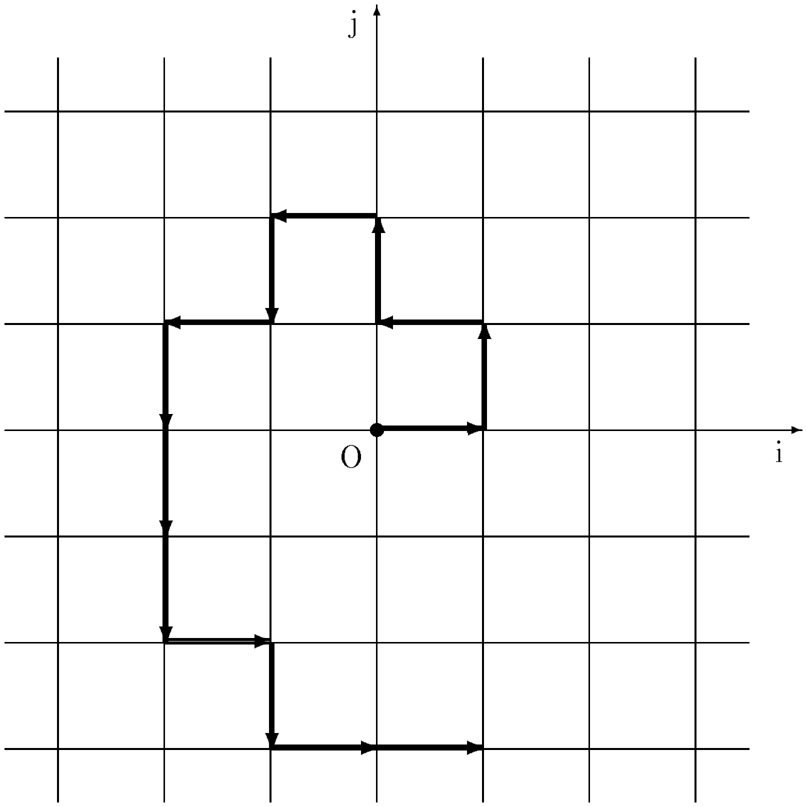

Figure 1.1 shows a finite part of an infinite graph called "checkered paper". Arrows show a finite path starting at O. This path has 14 steps and is self-avoiding because it never visits one and the same vertex twice. Generally, a path may visit any vertex many times.

Figure 1.1 represents an infinite graph, which we call "checkered paper". In mathematical terms, it is a graph, whose set of vertices is $\mathbb{Z}^2$, the set of pairs (i,j), where both i and j are integer numbers. Any vertex (i,j) is connected by non-oriented bonds with $(i+1,j)$, $(i,j+1)$, $(i-1,j)$. $(i,j-1)$. Let us imagine that every bond is a pipe which is either open or closed, namely it is open with probability ε and closed with probability $1-\varepsilon$ independently of other bonds. The origin $O=(0,0)$ is the only source of some liquid, which can pass along open bonds, but not along closed ones. There are no one-way bonds: if a bond is open, it is open in both directions. Vertices are always open. The vertices which the liquid can reach are called *wet*. The source is always wet by definition. For example, if the bond from $(0,0)$ to $(1,0)$ is open, the vertex $(1,0)$ is wet. If the bond from $(1,0)$ to $(1,1)$ is also open, the vertex $(1,1)$ is also wet and so on.

Let us call a *path* a sequence "vertex-bond-vertex-bond-..." in which every bond connects the vertices between which it is placed. If a path is finite, it ends with a vertex and the number of bonds in the sequence is called length of the path. One finite path of length 14 is shown on figure 1.1.

A path is called open if all its bonds are open. Clearly, a vertex is wet if there is an open path from the source to this vertex (or from this vertex to the source, which means the same). We say that *percolation* from O to ∞ occurs if the set of wet vertices is infinite. The most interesting feature of percolation is existence of a non-trivial critical value, which in the present case can be formulated as follows:

Theorem 1.1. Bond percolation from O to ∞ on the checkered paper has a critical value ε^* strictly between zero and one, such that:
a) If $\varepsilon < \varepsilon^*$, the probability of percolation is zero.
b) If $\varepsilon > \varepsilon^*$, the probability of percolation is positive.

In fact we shall prove that
a') if ε is small enough, the probability of percolation is zero and
b') if ε is large enough, the probability of percolation is positive.

This is sufficient to prove our theorem. Indeed, we can define ε^* as the supremum of those values of ε for which the probability of percolation is zero. According to what will be proved, thus defined ε^* is strictly between zero and one. Also it is easy to prove that the probability of percolation is a non-decreasing function of ε, whence the critical value is unique. It remains to prove a') and b'). In each case we have to start with some definitions and lemmas. A path is called *self-avoiding* if all the vertices in its sequence are different. To prove this theorem, we need the following lemma.

Lemma 1.1. Bond percolation from O to ∞ on the checkered paper is equivalent to existence of an open self-avoiding infinite path starting at the source O.

In one direction it is evident: if such a path exists, the set of its vertices is infinite and all of them are wet, so the set of wet vertices is also infinite.

Before arguing in the opposite direction, let us prove that for every wet vertex there is an open self-avoiding path from O to this vertex. Since this vertex is wet, there is some open path from O to it. If it is not self-avoiding, it visits some vertex twice and makes a loop between these visits. Let us eliminate this loop (including only one of these visits) from our path, thus obtaining a shorter path, which is still open and still leads from O to our vertex. If it is not yet self-avoiding, we eliminate a loop again an repeat this until we get a path without loops. Thus we get a self-avoiding open path from O to our vertex.

Now let us prove lemma 1.1 in the opposite direction. We can encode a path starting at O by the sequence of directions of its bonds as we pass them. For example, the path shown on figure 1.1 can be encoded as a sequence of directions *east, north, west, north, west, south, west, south, south, south, east, south, east, east.*

Let the set of wet vertices be infinite. Let us call S the set of open finite self-avoiding paths starting at O. Since for any wet vertex there is such a path leading there, S is infinite. Let us classify S into four subsets depending on direction of the first bond in the path:

$$S = S_{east} \cup S_{north} \cup S_{west} \cup S_{south}.$$

Since the union of these four sets is infinite, at least one of them is also infinite. Let it be S_{east} (the other cases are analogous). Then we classify S_{east} into three classes according to direction of the second bond:

$$S_{east} = S_{east,\ east} \cup S_{east,\ north} \cup S_{east,\ south}.$$

Again, at least one of these subsets must be infinite. If it is, say, $S_{east,\ north}$, we classify it again:

$$S_{east,\ north} = S_{east,\ north,\ east} \cup S_{east,\ north,\ north} \cup S_{east,\ north,\ west}$$

and again at least one of these subsets must be infinite. Thus we continue inductively. At the n-th step of our inductive argument we already have a sequence of n directions such that the set of open self-avoiding finite paths starting with these directions is infinite. Since we can continue this argument infinitely, this sequence grows infinitely, thereby defining an infinite open self-avoiding path starting at O. *Lemma 1.1 is proved.*

Now let us prove statement a'). If there is percolation, that is there is an infinite open self-avoiding path starting at zero, then, by taking its first n steps, we obtain a finite open self-avoiding path starting at O, whose length is n. Let us estimate the probability of its existence. For any self-avoiding path of lenght n the probability that it is open is ε^n. The number of self-avoiding paths of length n starting at O does not exceed $4 \cdot 3^{n-1}$. Thus the event "there is an open self-avoiding path of length n starting at O" is a union of at most $4 \cdot 3^{n-1}$ events, the probability of each being ε^n. Therefore the probability of this event does not exceed the sum of their probabilities, which is

$$4 \cdot 3^{n-1} \cdot \varepsilon^n \;=\; \frac{4}{3} \cdot (3\varepsilon)^n.$$

. If $\varepsilon < 1/3$, this quantity tends to zero when $n \to \infty$. But the probability of percolation is not greater than this quantity. Therefore the probability of percolation is zero for all $\varepsilon < 1/3$.

The proof of b') is more difficult. Which sets of closed bonds make percolation impossible? Let us call such sets *obstacles*. It is better to speak about minimal obstacles, that is obstacles, all of whose proper subsets are not obstacles. One minimal obstacle is shown on figure 1.2. Closed bonds are crossed and wet vertices are circled.

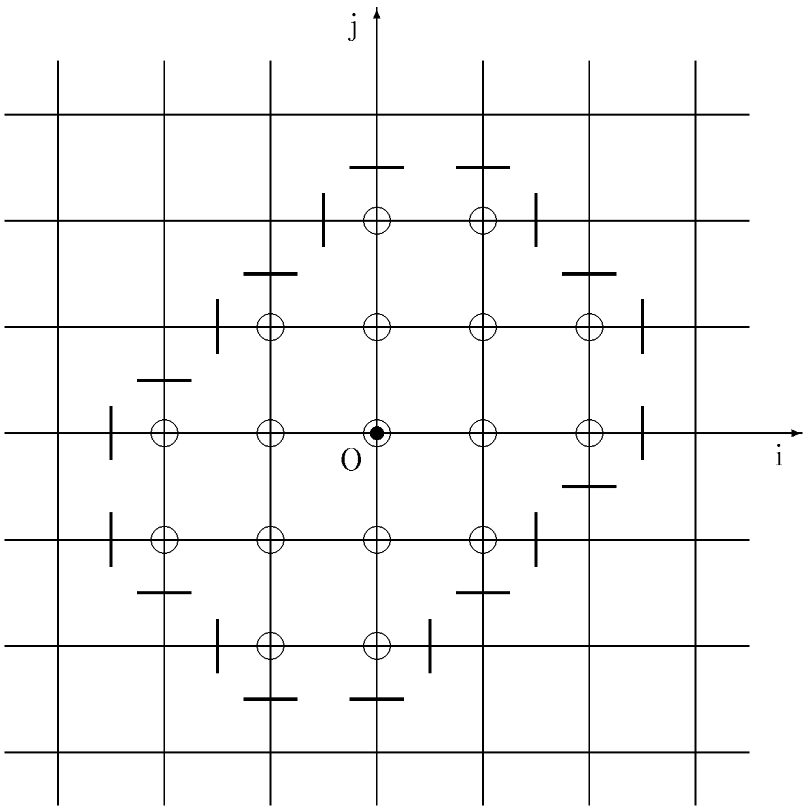

Figure 1.2. One minimal obstacle, that is a minimal set of closed bonds that makes percolation impossible. Closed bonds are crossed, wet vertices are circled.

You can see that closed bonds on figure 1.2 form some kind of fence around the origin. This becomes still more clear if we make the crossing bars longer, so that they form a continuous contour around O as shown on figure 1.3:

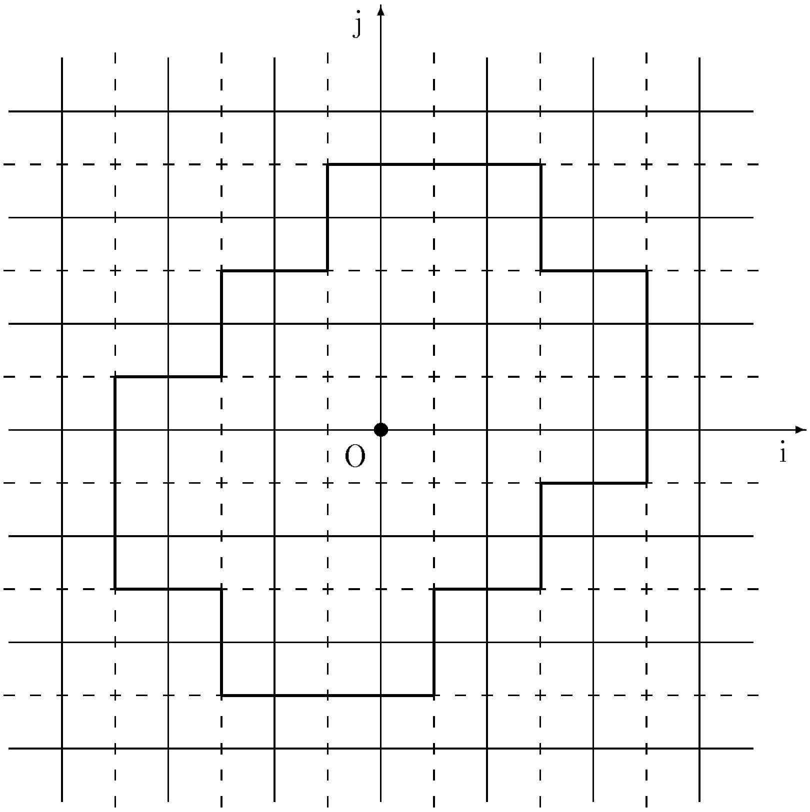

Figure 1.3. The crossing bars form a contour surrounding the origin.

This observation can be turned into a rigorous statement, but we must start with some definitions. Let us draw in the same plane another graph, which we call the *dual* graph. On figure 1.3 it is shown by dotted lines. There is a one-to-one correspondence between bonds of the two graphs, namely every bond of the dual graph crosses exactly one bond of the original graph and vice versa, and this is a relation between their being open:

Every bond of the dual graph is open if and only if the corresponding bond of the original graph is closed. (1)

Let us call a *contour* a self-avoiding path in the dual graph, whose starting and final points coincide. Every contour cuts the plane into two parts - finite and infinite and we say that a contour surrounds a point if this point is in the finite part. We call a contour open if all its bonds are open.

Lemma 1.2. If rule (1) is applied, there is no bond percolation in the checkered paper from O to ∞ if and only if in the dual graph there is an open contour surrounding O.

We leave proof of lemma 1.2 to the reader.

Now let us prove assertion b'). According to lemma 1.2, the probability that there is no percolation in checkered paper equals the probability of existence of an open contour surrounding O in the dual graph. This probability does not exceed the sum over all contours surrounding O of the probabililty that a given contour is open. Let us estimate this sum. All contours have an even number of steps and therefore this number can be denoted $2n$, where $n \geq 2$ because the minimal contour has length 4. A contour having $2n$ steps is open with a probability $(1-\varepsilon)^{2n}$. Thus the probability that there is no percolation does not exceed

$$\text{Prob(no percolation)} \leq \sum_{n=2}^{\infty} C_n\,(1-\varepsilon)^{2n},$$

where C_n is the number of different contours having $2n$ steps and surrounding O. It remains to estimate C_n. To determine a contour surrounding O and having $2n$ steps, it is sufficient:

i) Specify the i coordinate of the leftmost point of intersection of our contour with the positive half of axis i. This coordinate equals $k+1/2$, where k is an integer number between zero and $n-2$. (For example, $k=2$ on fugures 1.2 and 1.3.) Thus here we have $n-1$ cases.

ii) Specify directions of the $2n$ bonds starting from that which we hit in the item i) and going counter-clockwise along the contour. The first bond's direction is *north*, every other bond's direction has at most three possible values, the last bond's direction is predetermined because the contour must return to its initial point, so the number of cases here does not exceed 3^{2n-2}.

Therefore $C_n \leq (n-1)\cdot 3^{2n-2}$ and the probability that there is no percolation does not exceed

$$\sum_{n=2}^{\infty} (n-1)\cdot 3^{2n-2}\,(1-\varepsilon)^{2n}.$$

For $(1-\varepsilon)$ small enough this sum is less than one and this is what we need. In fact, this sum equals

$$\left(\frac{x}{3(1-x)}\right)^2 \quad \text{where} \quad x = \big(3(1-\varepsilon)\big)^2.$$

It is less than one if

$$\varepsilon \; > \; 1 - \frac{1}{2\sqrt{3}} \; \approx \; 0.71.$$

Thus the probability of percolation on checkered paper is zero if ε is small enough and positive if ε is large enough. We can define the *critical value* ε^* as the supremum of those values of ε for which the probability of percolation is zero. Then

$$0 \; < \; \frac{1}{3} \; \leq \; \varepsilon^* \; \leq \; 1 - \frac{1}{2\sqrt{3}} \; < \; 1.$$

Let us note an important property of percolation: monotonicity. It is evident that opening a bond can only promote percolation, but not hinder it. Threrefore the probability of percolation is a non-decreasing function of ε in all the cases considered in this chapter. Hence in every case there is only one critical value ε^* such that probability of percolation is zero for $\varepsilon < \varepsilon_*$ and positive for $\varepsilon > \varepsilon^*$. If thew critical value ε^* equals 0 or 1, we call it trivial; if it is in $(0,1)$, we call it non-trivial.

Thus we have proved our main statement: existence of a non-trivial critical value, which we define as the supremum of those values of ε for which the probability of percolation is zero.

Of course, our estimations of the critical value are very rough and can be improved. In fact, in the present case the critical value is known exactly: it equals $1/2$.[1] However, this is an exception connected with the fact that the dual graph of checkered paper is isomorphic with it. Dual graphs can be used in various cases, but generally they are not isomorphic with the original graphs and the critical values are not known exactly; we can only estimate them.

We formulated lemma 1.2 only for the checkered paper, but in fact it can be formulated in general terms, for any *planar graph*, that is a graph drawn on a plane so that its bonds do not intersect.

[1] To prove this is much more difficult. The latest comprehensive book on percolation is [Grimmett].

(For checkered paper the bonds are straight segments, but generally they may be curved.) For every planar graph there is a dual graph defined as follows. Let us call countries the parts into which the original graph cuts the plane. (For checkered paper countries are squares.) We choose a point in every country which we call its capital and these capitals are vertices of the dual graph. (For checkered paper the capitals are centers of the squares.) If two countries have a border, that is a common bond, the corresponding bond of the dual graph connects the capitals of these countries and crosses their common border. (For checkered paper these bonds are straight, but generally they may be curved.) Let us call O one vertex of the original graph and assume that it is the only source of liquid. As before, a vertex is wet if liquid can reach there and percolation means that the set of wet vertices is infinite.

Lemma 1.3. (A general version of Lemma 1.2.) If rule (1) is applied, there is no bond percolation from O to ∞ in a planar graph if and only if in its dual graph there is an open contour surrounding O.

We leave its proof to the reader. One of those graphs to which this statement allows to prove existence of critical value is shown on figure 1.4.

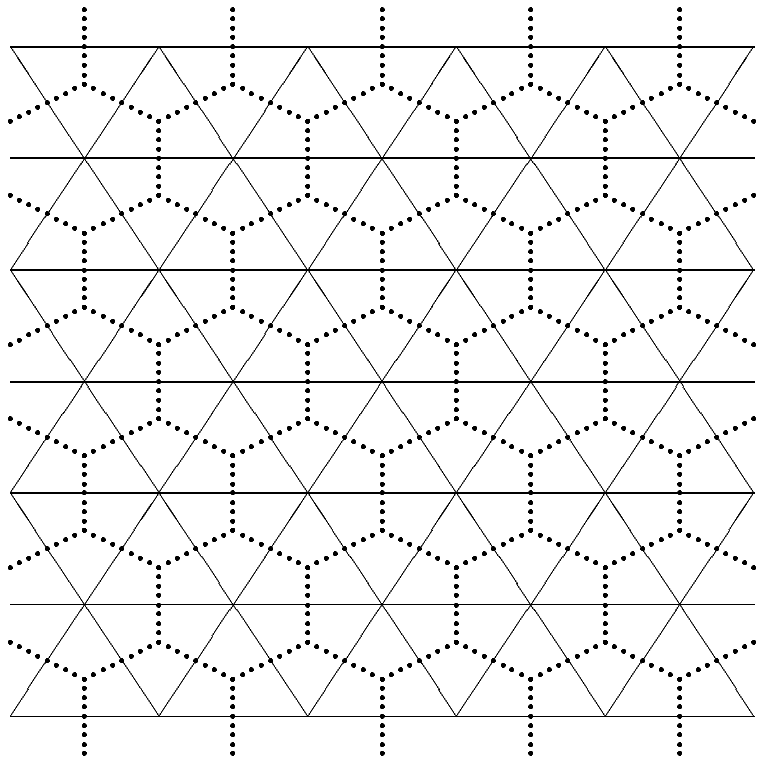

Figure 1.4. Continuous lines show a triangular lattice and dotted lines show the dual graph. In this case the dual graph is a hexagonal lattice.

Another kind of percolation, which is still more important in the present course, is oriented percolation. In this case bonds are oriented and we can pass them only in one direction. The figure 1.5 shows oriented analog of checkered paper.

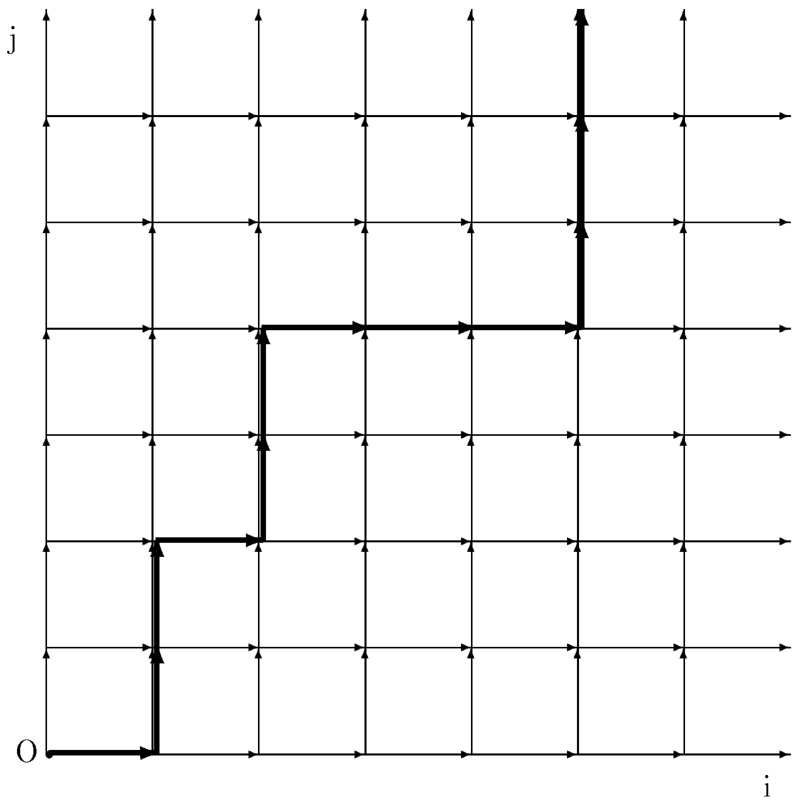

Figure 1.5. Oriented checkered paper. In this case a path can contain only east or north directed steps. One path starting at O is shown by arrows.

Suppose that every bond of figure 1.5 is open in the direction of arrow with probability ε and closed with probability $1-\varepsilon$ independently of others and always closed in the opposite direction. As before, we call a path open if all its steps are open - in that direction in which we use them - and a vertex is wet if there is an open path from O to this vertex. Percolation means, as before, that the set of open vertices is infinite or, which is equivalent, that there is an infinite open path starting at O.

Theorem 1.2. Bond percolation from O to ∞ on oriented checkered paper has a critical value ε^* strictly between zero and one such that:
a) If $\varepsilon < \varepsilon^*$, the probability of percolation is zero.
b) If $\varepsilon > \varepsilon^*$, the probability of percolation is positive.

As before, it is sufficient to prove that a') the probability of percolation is zero for ε small enough and b') the probability of percolation is positive for ε large enough.

The assertion a') is easy to prove. However, when proving the assertion b') we meet a new difficulty: the minimal obstacles are more complicated. Remember that our proof of ii) for theorem 1.1 was based on duality. Is there duality here?

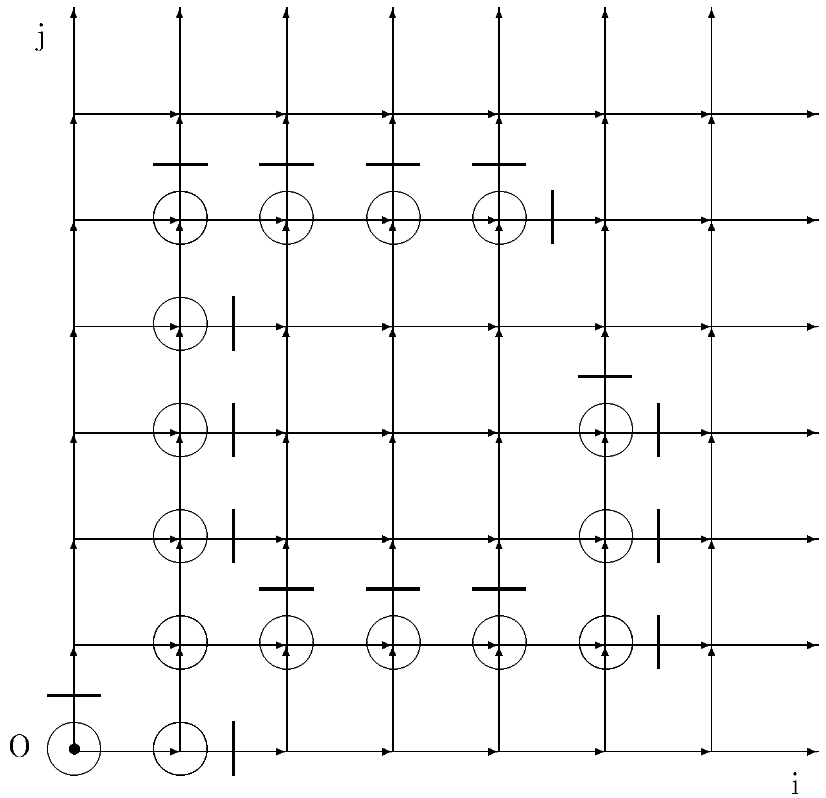

Figure 1.6. One minimal obstacle on oriented checkered paper. Wet vertices are circled. This obstacle also can be presented as a contour surrounding O. See next figure.

It turns out that the idea of duality works here also, but it is more complicated. First of all let us formulate how states of bonds of the dual graph depend on states of bonds of the original graph. Let us choose some orientation on the plane, namely counterclockwise. Then to every direction of a bond of the original graph there corresponds a direction of the corresponding bond of the dual graph: the direction from right to left.

$$\left.\begin{array}{l}\textit{Every bond of the dual graph is open in a certain direc-}\\ \textit{tion if and only if the corresponding bond of the original}\\ \textit{graph is closed in the corresponding direction.}\end{array}\right\} \quad (2)$$

Lemma 1.4. If rule (2) is applied, there is no bond percolation in a planar oriented graph from O to ∞ if and only if in the dual graph there is an open contour going around the source O in the positive direction, that is counterclockwise.

We leave proof of lemma 1.4 to the reader. It is illustrated by figure 1.7 where one contour is shown.

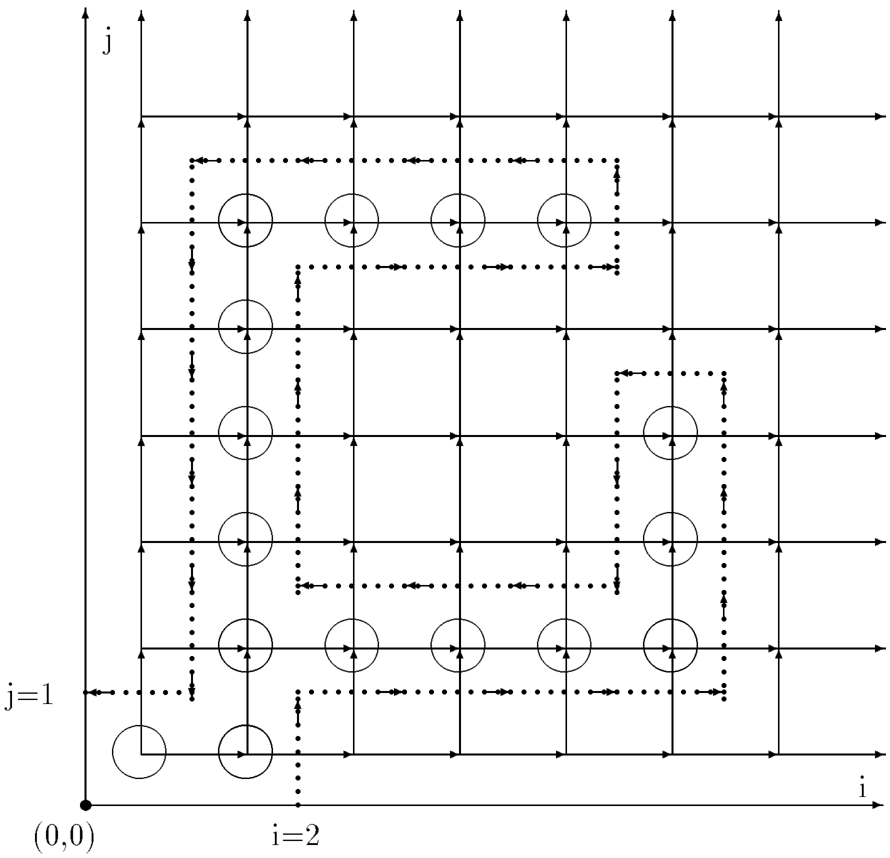

Figure 1.7. Contour corresponding to the obstacle shown on figure 1.6. Here $x = 2$ and $y = 1$ (coordinates of its beginning and end).

Let us use lemma 1.4 to prove the assertion b'). Like before, the probability that there is no percolation does not exceed

$$\sum_{\omega\in\Omega} \alpha^{|\omega|}, \tag{3}$$

where $\alpha = 1-\varepsilon$, Ω is the set of all minimal obstacles and $|\omega|$ is cardinality of ω. It follows from lemma 1.4 that to every minimal obstacle there corresponds a contour in the dual graph - like that shown on figure 1.7. The original graph cuts the plane into infinitely many squares and one unbounded country. Every contour in the dual graph starts and ends in one and the same country, namely in the unbounded one. Let us denote i the horizontal coordinate of its beginning and j the vertical coordinate of its end. For the contour shown on figure 1.7 $i=2$ and $j=1$ and generally i and j take any positive integer values. Also let us denote e, n, w, s the numbers of *east, north, west, south* steps in a contour. Notice that $w=i+e$ and $n=j+s$. The number of bonds in the corresponding obstacle is $n+w=i+j+e+s$. The total number of steps in the contour is $e+n+w+s=i+j+2e+2s$. The directions of the first and last steps are determined uniquely and directions of all the others are chosen of at most three options, so the number of contours with given e,n,w,s does not exceed $3^{e+n+w+s-2} = 3^{i+j+2e+2s-2}$. The table below shows the probabilities of original bonds to be open in all directions and probabilities of dual bonds resulting from the rule (2).

Original graph	*Dual graph*
east : open with prob. ε	*north*: open with prob. $\alpha = 1-\varepsilon$
north: open with prob. ε	*west* : open with prob. $\alpha = 1-\varepsilon$
west : always closed	*south*: always open
south: always closed	*east* : always open

Therefore the probability of an obstacle is $\alpha^{n+w} = \alpha^{i+j+e+s}$, whence (3) does not exceed

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{e=0}^{\infty}\sum_{s=0}^{\infty} 3^{i+j+2e+2s-2}\,\alpha^{i+j+e+s} =$$

$$\frac{1}{9}\cdot\sum_{i=1}^{\infty}(3\alpha)^i\cdot\sum_{j=1}^{\infty}(3\alpha)^j\cdot\sum_{e=0}^{\infty}(9\alpha)^e\cdot\sum_{s=0}^{\infty}(9\alpha)^s =$$

$$\frac{1}{9}\cdot\frac{3\theta}{1-3\alpha}\cdot\frac{3\theta}{1-3\alpha}\cdot\frac{1}{1-9\alpha}\cdot\frac{1}{1-9\alpha} =$$

$$\left(\frac{\alpha}{(1-3\alpha)(1-9\alpha)}\right)^2.$$

If α is small enough, e.g. smaller than 0.09, this expression is less than 1. Thus we have proved that the probability of percolation is positive if $\varepsilon > 1 - 0.09 = 0.91$. This is only an estimation. The critical value is less than this; its exact value is unknown.

There are other kinds of percolation. One of them is especially important for us. In this case we have a finite planar oriented graph in which bonds may be open in one direction and closed in the other. A path is defined as before, but now it has a direction and is considered open if all its bonds are open in its direction. Percolation from vertex A to vertex B means existence of an open path from A to B. The dual graph is defined as before with the same rule (2) defining how openness of its bonds depends on the openness of bonds of the original graph.

Lemma 1.5. If the rule (2) is applied, there is no bond percolation from a vertex A to another vertex B in a planar oriented graph if and only if in the dual graph there is an open contour separating A from B and going around B in the positive direction.

This lemma will be used in the next chapter. We leave its proof to the reader.

Notes.

1.1. *Exercise.* What does lemma 1.3 mean if the graph is finite?

1.2. *Exercise.* Prove that the probability of percolation is a non-decreasing function of ε for all the cases mentioned in this chapter.

1.3. *Exercise.* Prove existence of critical value for the triangular lattice shown on figure 1.4.

1.4. *Exercise.* Prove lemmas 1.2, 1.3 and 1.4.

1.5. *Exercise.* Let us consider one-dimensional percolation, where the set of vertices is $\mathbb{Z}$ and two vertices x and y are connected with a bond if $|x-y| \le 100$. As before, suppose that 0 is the only source of liquid, every bond is open with a probability ε independently from other bonds and percolation means that the set of wet vertices is infinite. Prove that in this case the probability of percolation is zero for all $\varepsilon < 1$. As before, we can define ε^*

as the supremum of those values of ε for which the probability of percolation is zero, but now it is trivial: $\varepsilon^* = 1$. This suggests that there is qualitative difference between one-dimensional and multi-dimensional cases in percolation.

1.6. *Exercise.* What we described in this chapter was *bond* percolation because only bonds could be closed. Another kind of percolation is *vertex percolation* when vertices may be open or closed. Let us consider just one case, which is close to our interests: *vertex percolation* on oriented checkered paper shown on figure 1.5. As before, the origin O is the only source of liquid, but now bonds are always open in the directions of arrows and always closed in the opposite directions. Every vertex is open with probability ε and closed with probability $1-\varepsilon$ independently of other vertices. A path is defined as before, but now it is open if all its vertices are open. As before, a vertex is called wet if there is an open path from O to this vertex and percolation means that the set of wet vertices is infinite. In this case also there is critical value ε^* strictly between zero and one such that:
a) If $\varepsilon < \varepsilon^*$, the probability of percolation is zero.
b) If $\varepsilon > \varepsilon^*$, the probability of percolation is positive.
Prove this and show that any minimal obstacle can be represented as a contour in some graph. The value of this critical value is not known exactly; we can only estimate it.

1.7. *Solved problem.* Probability of percolation is a continuous function of ε in all the cases considered here [Grimmett]. This is especially difficult to prove at the critical point.

1.8. *Unsolved problem.* How to define duality for dimensions higher than two?

2. Some cellular automata are similar to percolation

Let us define a class of cellular automata about which we shall speak in this chapter. The space where components are placed is $U = \mathbb{Z}^d$, the d-dimensional integer space. Elements of U are called points or sites. We imagine that every point $i \in U$ can either be empty or contain a particle. All particles are undistinguishable and a point cannot contain more than one particle. Thus every point has only two possible states, which we denote 0 and 1, where 0 means an empty point and 1 means that there is a particle there, whence the configuration space is $\Omega = \{0,1\}^U$. Any configuration $x \in \Omega$ is given by its components $x_v \in \{0,1\}$, $v \in U$. For any configuration x we denote $I_0(x)$ the set of points v where $x_v = 0$ and: $I_1(x)$ the set of points v where $x_v = 1$. For any finite list $i_1, \ldots, i_n \in U$ and any $a_{i_1}, \ldots, a_{i_n} \in \{0,1\}$ there is a subset of Ω of the form

$$\{x \in \Omega \ : \ x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}\}. \tag{4}$$

We call such sets *thin cylinders.*[2] The set $\{i_1, \ldots, i_n\}$ is called *support* of this thin cylinder. Let us use the following operations: taking complement and taking finite or countable union or intersection. All the subsets of Ω which result from several applications of these operations to thin cylinders form a σ-algebra (or Borel field) of subsets of Ω. We call these subsets *measurable* and we denote $\mathcal{M}$ the set of probability distributions, i.e. normed measures on this σ-algebra. For short, we call them measures on Ω and denote by μ and other Greek letters. The word "normed" means that $\mu(\Omega) = 1$; often it will be omitted, because we consider only normed measures. Values of a measure on thin cylinders must be consistent in the following sense: for any $i_0, i_1, \ldots, i_n$

$$\sum_{a_{i_0}} \mu(x_{i_0} = a_{i_0}, x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}) =$$

$$\mu(x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}). \tag{5}$$

Pay attention to our simplified notation. For example, (5) should be, formally speaking, be written as

$$\mu\left(\{x \in \Omega \ : \ x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}\} \right),$$

but we allow ourselves to abbreviate whenever it does not cause confusion. If a measure equals zero at at least one thin cylinder, we call it *degenerate*, otherwise it is non-degenerate. If a measure is concentrated in one configuration x, we call it a *delta-measure* and

[2] The word "cylinders" is kept for finite unions of thin cylinders.

denote $\delta(x)$. Also we use special abbreviations $\delta_0 = \delta(\text{all zeros})$ and $\delta_1 = \delta(\text{all ones})$. Let us say that a sequence of normed measures on Ω converges if it converges on all thin cylinders.

Lemma 2.1. If a sequence of normed measures on Ω converges, it converges to a normed measure on Ω.

Proof. We need only to check that (5) is fulfilled for the limit. This is evident since the summation is finite.

Generally speaking, if two measures coincide on all thin cylinders, whose support consists of one element, these still may be different. However, if one of these measures is a delta-measure, the measures have to coincide. The following lemma says this and more.

Lemma 2.2.
Given a configuration $a = (a_v) \in \Omega$.
a) If $\mu(x_v = a_v) = 1$ for all $v \in U$. then the measure μ is a delta-measure $\delta(a)$ concentrated in the configuration $a = (a_v)$.
b) If there is a sequence of measures $\mu_1, \mu_2, \mu_3, \ldots$ such that

$$\lim_{n\to\infty} \mu_n(x_v = a_v) = 1 \quad \text{for all} \quad v \in U$$

then μ_n tend to $\delta(a)$ when $n \to \infty$.

Proof is left to the reader.

Now let us speak about cellular automata. All of them are linear operators acting on $\mathcal{M}$, so we shall often use the word "operator" instead of a longer phrase "cellular automaton". The word "process" means a sequence of measures μ, $P\mu$, $P^2\mu$, ... resulting from iterative application of an operator P to some initial measure μ. A measure μ is called *invariant* for P if $P\mu = \mu$. We call an operator $P : \mathcal{M} \to \mathcal{M}$ *ergodic* if the limit $\lim_{t\to\infty} P^t\mu$ exists and is one and the same for all μ. If P is ergodic, it has only one invariant measure, but the converse is not proved. In the last chapter we shall prove that all cellular automata of a large class have at least one invariant measure. Generally, it is important to study ergodicity and sets of invariant measures of cellullar automata.

Ergodic operators correspond to real systems which "forget" their initial conditions - this is what we usually want to achieve when we mix a drink. Non-ergodic operators correspond to real systems which partially remember their initial conditions - this is what we want to achieve when we keep information in computer memory.

The central goal of this course is to present some examples of ergodic and some examples of non-ergodic cellular automata.

Now let us present our first examples of cellular automata. Let the particles do the two basic functions of living beings: reproduce and die and let them reproduce deterministically and die randomly. So our operator is a superposition $P = R_\alpha\, D$, which means that action of P consists of two steps: first deterministic reproduction D, then random death R_α, where $\alpha \in [0,1]$.

Operator D is deterministic. This means that it acts on configurations: $D : \Omega \to \Omega$. To define D, we need to choose a non-empty finite list of vectors $v_1, \ldots, v_n \in \mathbb{Z}^d$, which we call *neighbor vectors*. Points $v + v_1, \ldots, v + v_n$ are called neighbors of v. Under the action of D every particle, which is at some point v, generates particles in all the points $v - v_1, \ldots, v - v_n$ unless these points are already occupied. In other words, there will be a particle at the point v after application of D if and only if there was a particle in at least one of the points $v + v_1, \ldots, v + v_n$ before its application. In mathematical terms, D transforms any configuration x into a configuration $D\,x$, whose i-th coordinate is

$$(D\,x)_v \;=\; \max\,(x_{v+v_1}, \ldots, x_{v+v_n}) \quad \text{for all} \quad v \in \mathbb{Z}^d. \qquad (6)$$

Now let us define R_α, random death. Under its action every particle dies (1 turns into 0) with probability α independently of what happens to other particles. Thus $\alpha \in [0,1]$, called *death rate*, is the only parameter and we shall show that there is a non-trivial critical value α^*. We call superpositions $P = R_\alpha\, D$ "percolation operators", you will see why. Percolation operators are the simplest examples of cellular automata showing the interesting and important phenomenon of phase transition.

Notice that our operators commutate with shifts of the space. We shall call such operators *uniform*. *Uniform measures* are those invariant under space shifts. Dealing with a uniform measure, we may speak of density of ones or density of zeros. If a uniform operator P is applied to a uniform initial measure, the resulting measures are uniform also. Also, if a deterministic operator is uniform, it is sufficient to specify how it transforms the O-th component; this determines how it transforms all the others. Throughout chapters 2,3,4 operators and measures are uniform.

Now let us look what happens if D or R_α acts alone, without the other. In these cases the situation is easy to examine. If D is iterated alone and there are some particles scattered randomly at the beginning, they reproduce freely (provided $n \geq 2$) and as time (number of iterations) tends to ∞, they fill all the space. However, if initially there are no particles, that is if the initial condition is "all zeros", the process remains there forever. If R_α is iterated alone, then, whatever is the initial condition, particles die out and their density tends to zero as time tends to ∞. Now let us consider their superposition $P = R_\alpha D$, which means that at every time step first D acts, then R_α acts. What will happen to the density of particles as time goes to infinity?

Theorem 2.1. Every percolation operator $P = R_\alpha D$ with $n \geq 2$ has a non-trivial critical value $\alpha^* \in (0, 1)$ such that:
a) if $\alpha > \alpha^*$, the particles die out from any initial condition.
b) if $\alpha < \alpha^*$ and initially all the space is filled with particles, particles do not die out.
c) $1/54 \leq \alpha^* \leq 1 - 1/n$.

In the case a) P is ergodic because the process converges to the distribution concentrated in the configuration "all zeros" from any initial condition. In the case b) P is not ergodic because if the process starts from "all zeros", it remains there forever, but if it starts from "all ones", it does not tend to "all zeros".

Let us prove theorem 2.1 in one direction: if α is large enough, particles die out. Notice that any percolation process can be represented using oriented percolation (that is why we call it this way). Let us consider *percolation graph* with vertices denoted (v, t), where $v \in \mathbb{Z}^d$ and $t = 0, 1, 2, 3, \ldots$. From every vertex (v, t) oriented bonds go to vertices $(v - v_1, t+1), \ldots, (v - v_n, t+1)$. Every bond is open in the direction of its orientation (from t to $t+1$) and closed in the opposite direction. Let us also assume that every vertex can be either open or closed. All the initial vertices $(v, 0)$ are open (this means that the initial condition is "all ones"), all the other vertices are closed with probability α and open with probability $1 - \alpha$ independently of each other. A path is defined as in chapter 1 and it is open if all its vertices are open.

Lemma 2.3. In a percolation process starting from "all ones" there is a particle at a point v at time t if and only if there is an open path from some initial vertex to the vertex (v, t) in the percolation graph.

We leave proof of lemma 2.3 to the reader. Let us use it to prove that if $\alpha > 1 - 1/n$, particles die out. Indeed, for any path from any initial vertex to a vertex (v, t) the probability that it is open is $(1-\alpha)^t$ and there are n^t such paths. Therefore the probability that there is a particle at a certain point at time t does not exceed

$$(1-\alpha)^t \cdot n^t \;=\; ((1-\alpha)n)^t.$$

If $\alpha > 1 - 1/n$, this quantity tends to zero when $t \to \infty$, whence the particles die out. This proves theorem 2.1 in one direction. We can define α^* as the infimum of those α, for which particles die out from any initial condition. We have proved the item c): $\alpha^* \le 1 - 1/n$.

Now let us prove theorem 2.1 in the opposite direction: if α is small enough and all the space was filled with particles in the beginning, they do not die out. In this chapter we shall prove this only for *Stavskaya operator*,[3] which is the simplest percolation operator: here $d = 1$ and there are only two neighbor vectors: $0, 1 \in \mathbb{Z}$. Notice that presence of a particle at a point i at time t depends only on what happens in the triangle

$$\{(j,s) \;:\; i \le j \le i + t - s\}.$$

Figure 2.1 shows this triangle for $i = 0$ and $t = 3$. (The axis of time is slanted to make the scheme symmetric.)

$$\begin{array}{ccccccc}
 & & & (0,3) & & & \\
 & & & \nearrow \quad \nwarrow & & & \\
 & & (0,2) & & (1,2) & & \\
 & & \nearrow \quad \nwarrow & & \nearrow \quad \nwarrow & & \\
 & (0,1) & & (1,1) & & (2,1) & \\
 & \nearrow \quad \nwarrow & & \nearrow \quad \nwarrow & & \nearrow \quad \nwarrow & \\
(0,0) & & (1,0) & & (2,0) & & (3,0)
\end{array}$$

Figure 2.1. The triangle on which the state of point 0 *at time* 3 *depends in Stavskaya process.*

The figure 2.2 shows that part of the percolation graph for Stavskaya process, which is relevant for the state of point 0 at time 3:

[3] What we denote 0 was denoted 1 and vice versa in [Sta+Pia] and [Discr], because initially the Stavskaya operator was introduced as a model of spontaneously-active formal neurons.

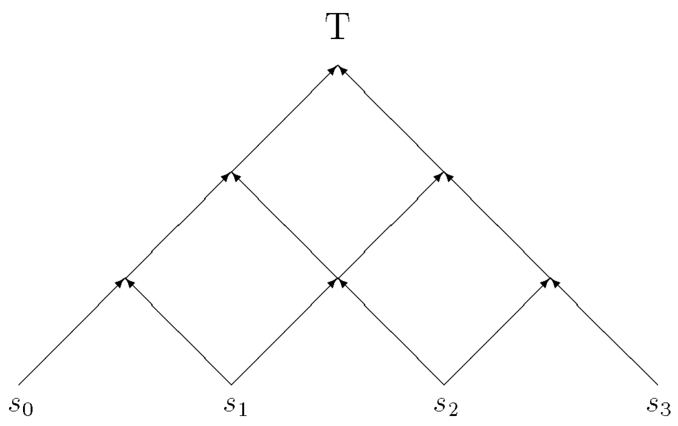

Figure 2.2. Part of percolation graph for Stavskaya process. Bonds are always open upward and closed downward. Vertices s_0, s_1, s_2, s_3 are always open. Other vertices are open with probability $1-\alpha$ and closed with probability α independently of each other. There is a particle at the point 0 at time 3 if and only if there is an open path in this graph from some of the sources s_i to the target T.

However, it is better to stretch every vertex, thus turning vertex percolation into bond percolation as shown on figure 2.3.

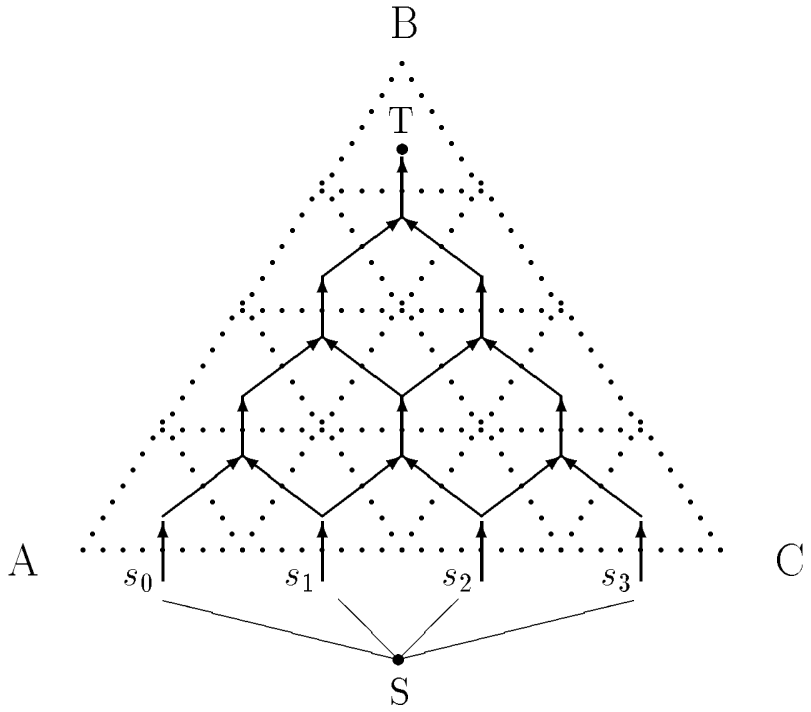

Figure 2.3. Stavskaya process as bond percolation. Presence of a particle at the point $(0,3)$ *amounts to percolation in the graph shown by contunuous lines from the source* S *to the target* T*. Dotted lines show the dual graph. The sides* AB *and* BC *correspond to one country.*

Let us imagine that four vertices denoted s_0, s_1, s_2, s_3 on figure 2.3 are sources of liquid and that arrows are oriented pipes which can transmit this liquid upward, but not back. The inclined arrows are always open, but the vertical arrows may be open or closed because they imitate our random operator: everyone of them is closed with probability α and open with probability $1-\alpha$. Then the probability that there is a particle at point 0 at time 3 in our process equals the probability that there ia an open path from at least one of the sources s_0, s_1, s_2, s_3 to the target T. Thus we have reduced a problem about our random process to a problem about percolation.

However, it is better to have only one source. For this reason we introduce a special vertex S and connect in by bonds with s_0, s_1, s_2, s_3. It is convenient to assume that these bonds are always open in both directions - then the dual bonds will be always closed in both directions and we don't even need to draw them. For the same reason it is convenient to assume that the vertical bonds of the original graph are always open downward; it does not create any unwanted opportunities of percolation because the slanted bonds are always closed downward. Then, according to the rule (2), the bonds of the dual graph (shown by dotted lines) are open as follows: bonds directed $\swarrow$ and $\nwarrow$ are always open in these directions and always closed in the opposite directions; bonds directed $\rightarrow$ are open in this direction with probability α and always closed in the opposite direction.

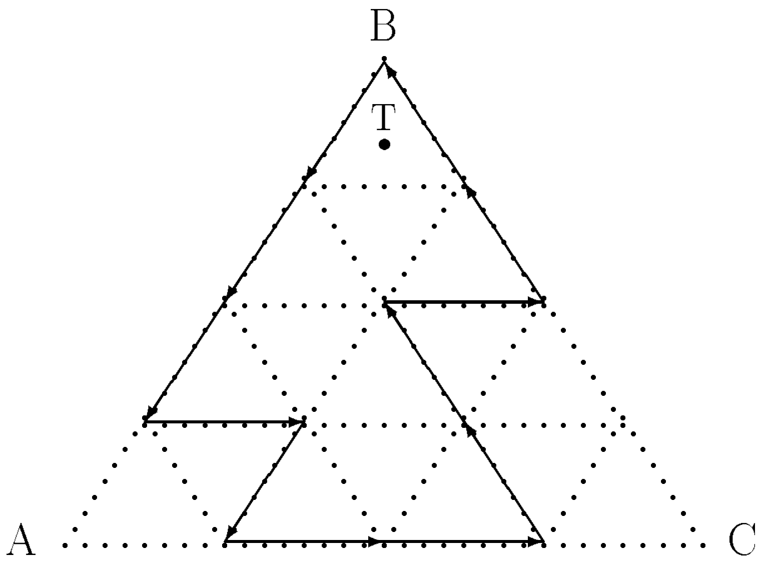

Figure 2.4. The dual graph for the Stavskaya process. Continuous arrows show one contour surrounding T. Existence of an open contour surrounding T in the positive direction amounts to absence of particle at the point 0 at time 3 in the process.

Let us concentrate our attention at the dual graph shown on figure 2.4. According to lemma 1.4, there is no percolation in the original graph if there is an open contour in this graph surrounding T and going in the positive (counterclockwise) direction. We may assume that every contour starts and ends at the topmost point B. The probability that there is such a contour does not exceed

$$\sum_{k=1}^{\infty} C_k\, \alpha^k, \tag{7}$$

where C_k is the number of such contours corresponding to obstacles with k elements, that is having k horizontal steps. Every contour has equal numbers of steps of all the three directions, so altogether it has $3k$ steps. Since every step of a contour has only three possible directions, $C_k \leq 3^{3k}$ and therefore the probability that there is a particle at site 0 at time t does not exceed

$$\sum_{k=1}^{\infty} 3^{3k} \cdot \alpha^k = \frac{27\alpha}{1-27\alpha}, \tag{8}$$

which is less than one as soon as $\alpha < 1/54$. Thus, whenever $\alpha < 1/54$, particles do not die out because their density does not tend to zero. We have proved that $1/54 \leq \alpha^* \leq 1/2$ for Stavskaya operator.

We have not yet proved theorem 2.1 for all percolation operators. If $n = 2$, it is easy because of the following lemma:

Lemma 2.4. The density of particles at any time t is one and the same for all percolation operators with $n = 2$ if the initial condition is "all ones".

Proof is left to the reader. The remaining part of the proof of theorem 2.1 is in the last chapter.

Suppose we want to model Stavskaya process on a computer. The most usual method to do it is Monte Carlo. The following pseudo-code shows how to do it, where the lines are enumerated for reference purpose:

1 for all $i \in \mathbb{Z}_m$ do
2 $\quad x_{i,\,0} \leftarrow 1$
3 for $t = 1$ to t_{max} do
4 $\quad$ for all $i \in \mathbb{Z}_m$ do
5 $\qquad x_{i,\,t} \leftarrow \max(x_{i,\,t-1},\ x_{i+1,\,t-1})$
6 $\qquad$ if $rnd < \alpha$ then $x_{i,\,t} \leftarrow 0$

Here the space is the set $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$ of residues modulo m.[4] The sign $\leftarrow$ in line 2 is the assignment operator; $x \leftarrow a$ means that variable x is assigned value a. Thus lines 1-2 assign the initial configuration "all ones". Lines 4-6 perform a time step. Line 5 corresponds to the deterministic operator. Line 6 corresponds to the random operator. It uses a random number rnd which is uniformly distributed between 0 and 1, newly generated every time when it is called and is independent from all the previous random numbers.

Of course, in computer simulation the space has to be finite and time cannot grow to infinity; it grows to some value t_{max} when computation stops. However, in our imagination the space and time span may be infinite:

1 for all $i \in \mathbb{Z}$ do
2 $\quad x_{i,\,0} \leftarrow 1$
3 for $t = 1$ to ∞ do
4 $\quad$ for all $i \in \mathbb{Z}$ do
5 $\qquad x_{i,\,t} \leftarrow \max(x_{i,\,t-1}, x_{i+1,\,t-1})$
6 $\qquad$ if $rnd < \alpha$ then $x_{i,\,t} \leftarrow 0$

[4]Operations with resudues modulo m result in residues modulo m, in particluar $(m-1) + 1 = 0$.

This pseudo-code is similar to the former one, only the space is $\mathbb{Z}$ and the time span is from 1 to ∞. It cannot be used for actual programming, because computer memory and time are always finite, but we shall use pseudo-codes of this sort to define some processes.

In mathematical terms, the Stavskaya process is *induced* by independent random variables $y(i,t)$, everyone of which equals 0 with probability α and 1 with probability $1-\alpha$ by the map defined in the following inductive way:

Base of induction: $x(i,\ 0) = 1$ for all $i \in \mathbb{Z}$.

Induction step: $x(i,\ t) = y(i,\ t) \cdot \max(x(i,\ t-1),\ x(i+1,\ t-1))$.

The Stavskaya process was modelled on a computer and the observed results supported theory: if α was small, the particles survived, if α was large, they died out. But there seems to be a contradiction here, because only finite processes can be modelled on computer, but it is easy to prove that for any finite process the particles die out in a finite mean time for any $\alpha > 0$. To prove this is easy, it is sufficient to notice that all the particles may die at once with a probability α^m or even greater. Hence the expectation of time when all particles die out is not greater than α^{-m}. Why this was not observed in computer experiments? Notice that the number α^{-m} is enormous even for quite moderate values of α and m, much greater that the time which we can afford in experiment. What we observe in the experiment is not distinction between ergodicity and non-ergodicity which takes place in infinite processes, but distinction between fast and slow convergence which takes place in finite processes. This distinction can be formulated as follows.

Theorem 2.2 [Discr]. In finite Stavskaya processes with initial condition "all ones":
a) If $\alpha \in (0,1)$ is small enough, the mathematical expectation of time, when all particles die out, grows as an exponent of m when $m \to \infty$.
b) If α is large enough, the mathematical expectation of time, when all particles die out, grows as a logarithm of m when $m \to \infty$.

This theorem can be proved using the same ideas as theorem 2.1. Thus for finite processes also there is some kind of phase transition. See below an unsolved problem in this connection.

The next theorem provides us with a rich supply of ergodic operators. This time we use a much more general class of deterministic operators. To define a deterministic operator D, we take a non-empty finite list of vectors $v_1, \ldots, v_n \in \mathbb{Z}^d$ and a Boolean function $f : \{0,1\}^n \to \{0,1\}$. Then D is defined as follows:

$$(Dx)_v \;=\; f(x_{v+v_1}, \ldots, x_{v+v_n}) \quad \text{for all} \quad v \in \mathbb{Z}^d. \tag{9}$$

Also we use a random operator R_α^β, which turns every 1 into 0 with probability α and turns every 0 into 1 with probability β and it does all these changes independently of each other. Of course, $R_\alpha = R_\alpha^0$, so this operator is a generalization of R_α.

Theorem 2.3. Let D be defined by (9) and

$$1 - \frac{1}{n} \;<\; \alpha + \beta \;\leq\; 1.$$

Then the operator $R_\alpha^\beta\, D$ is ergodic.

The main notion, which we use to prove this theorem is *coupling* of several processes, that is a process on a space which a product of their spaces, whose marginals are given processes. In the present case we use a coupling of three processes: two processes generated by our operator $R_\alpha^\beta\, D$ with different initial conditions and a percolation process. This coupling is defined by the following pseudo-code, where $x(v,t)$, $y(v,t)$ and $m(v,t)$ mean components of the three marginals at a point v at time t:

```
for all v ∈ Z^d do
   x(v,0) ← x_initial(v,0)
   y(v,0) ← y_initial(v,0)
   m(v,0) ← 1
for t = 1 to ∞ do
   for all v ∈ Z^d do
      x(v,t) ← f(x(v+v_1,t−1),...,x(v+v_n,t−1))
      y(v,t) ← f(y(v+v_1,t−1),...,y(v+v_n,t−1))
      m(v,t) ← max(m(v+v_1,t−1),...,m(v+v_n,t−1))
      r ← rnd
      if r < α then
         x(v,t) ← 0
         y(v,t) ← 0
         m(v,t) ← 0
      if r > 1 − β then
         x(v,t) ← 1
         y(v,t) ← 1
         m(v,t) ← 0
```

Let us first ignore all the lines of this pseudo-code which deal with the values $m(v,t)$ and concentrate our attention on those, which deal with $x(v,t)$ and $y(v,t)$. They describe two processes with one and the same operator $R_\alpha^\beta\ D$ and arbitrary different initial conditions set by lines 1-3. These processes function simultaneously, using a common source of random noise. In both processes every component every time does the following: first, due to lines 7 and 8, it assumes some value, which depends on states of its neighbors one time step ago, and second, due to lines 10-17 it makes a random change, becoming 0 with probability α and becoming 1 with probability β. Let us call a point (v,t) a *point of difference* if $x(v,t) \neq y(v,t)$.

Lemma 2.5. Suppose that points of difference in the coupling defined by this pseudo-code die out uniformly in v and initial condition. In other words, suppose that there is a sequence p_t, which tends to 0 when $t \to \infty$, such that for any v and any initial condition the probability that $x(v,t) \neq y(v,t)$ does not exceed p_t. Then operator $P = R_\alpha^\beta\ D$ is ergodic.

We shall prove this lemma in the last chapter. Now let us check that points of difference really die out uniformly. To monitor what happens to points of difference, we have special *marks* $m(v,t)$, which may equal 0 or 1. We call a point (v,t) *marked* if $m(v,t)=1$ and unmarked otherwise. The interaction is arranged in such a way that $m(v,t)=1$ at every point of difference:

$$\forall\ v,\ t\ :\ x(v,t) \neq y(v,t) \implies m(v,t)=1.$$

(The opposite implication may be false, that is $m(v,t)$ may equal 1 at other points also.) Initially all the points are marked (to be on the safe side, because initially all points may be points of difference) and become unmarked only when lines 12-13 or 16-17 assign equal values to $x(v,t)$ and $y(v,t)$. Notice that the mark process is a percolation process with the death rate $\alpha+\beta$. Therefore, from the item c) of theorem 2.1, marked points die out whenever

$$1-\frac{1}{n}\ <\ \alpha+\beta\ \leq\ 1.$$

Theorem 2.3 is proved except lemma 2.5.

Notes.

2.1. *Exercise.* Prove that all percolation operators with the number of neighbors $n=1$ are ergodic for all positive α.

2.2. *Exercise.* Prove that density of particles at any time t in Stavskaya process is a non-increasing function of α.

2.3. *Exercise.* Let us take a percolation operator P, apply it t times to the initial configuration containing only one particle and denote $p_{extint}(t)$ the probability that after t steps there are no more particles. Also let us apply P to the initial condition "all space is filled with particles" and denote $p_{empty}(t)$ the probability that the 0-th site is empty after t steps. Prove that $p_{extint}(t) = p_{empty}(t)$.

2.4. *Exercise.* Prove that as soon as the series (7) converges, the Stavskaya operator is non-ergodic and use this to prove that $\alpha^* \geq 1/27$.

2.5. *Solved problem.* Notice that a linear operator acting on $\mathcal{M}$ cannot have exactly two invariant measures: as soon as two different measures μ and ν are invariant for it, all their linear combinations $k\mu + (1-k)\nu$ for $0 < k < 1$ also are. So, when we study the set of invariant measures of an operator, we should ask, how many linearly independent invariant measures it has. In particular, it is proved that the Stavskaya operator P has at most two linearly independent invariant measures: one is δ_1, the other is the limit of $P^t\delta_0$ when $t \to \infty$. These two measures coincide if $\alpha \geq \alpha^*$ and are different if $\alpha < \alpha^*$.

2.6. *Solved problem.* Let us change the law of reproduction in Stavskaya process as follows: every particle, in addition to itself, produces a new particle either on its left side or on its right side at random (whenever this place is empty). It has been proved that theorem 2.1 is true in this case also. (It follows from the theory of random walk operators [Discr].)

2.7. *Solved problem.* Let us rename zeros into ones and vice versa in the Stavskaya process and change our interpretation to the opposite: now the deterministic operator is death of a particle and the random operator is birth of a particle and let it be possible for a point to contain many particles. The initial configuration is "all zeros" and at every time step two operators act: first deterministic D, then random R_α. Under the action of D every component goes into state which is the minimum of its state and its right neighbor's state. Under the action of R_α every component increases its state by one with probability α and remains in the same state with probability $1-\alpha$ independently of what happens to other components. In this case also there is a critical value α^*_{growth} such that:
a) if $\alpha < \alpha^*_{growth}$, the mathematical expectation of a component's

state tends to a finite limit when $t \to \infty$.
b) if $\alpha > \alpha^*_{growth}$, the mathematical expectation of a component's state tends to ∞ when $t \to \infty$.
This is a special case of a more general theorem proved in [T-94]. However, you can prove it without looking there using ideas of our chapter 2.
Unsolved problem: Are the critical value for Stavskaya operator and the critical value for the growth Stavskaya operator equal?

2.8. *Unsolved problem.* Let us define two critical values for finite Stavskaya operators: α^*_{exp} the supremum of those α for which case a) of theorem 2.2 takes place and α^*_{log} the infimum of those α for which case b) of theorem 2.2 takes place. It is evident that $\alpha^*_{exp} \leq \alpha^*_{log}$ and it is proved that $\alpha^* \leq \alpha^*_{log}$, but is not known whether these three quantities are equal. See more about it in [Cell], p. 140.

2.9. *Unsolved problem. Vasilyev operator* is similar to Stavskaya operator, but it is non-monotonic (monotonicity is introduced in the last chapter), which makes it much more difficult to study. Its most interesting property, existence of a critical value is not yet proved, although it is strongly suggested by computer simulation. Its operator also is a superposition of two operators: deterministic and random. *Deterministic operator* D is defined by $(Dx)_i = x_i \oplus x_{i+1}$, where the binary operation $\oplus$ is sum modulo 2, defined as follows:

$$0 \oplus 0 = 1 \oplus 1 = 0,\ 0 \oplus 1 = 1 \oplus 0 = 0.$$

Random operator R_α is the same as in Stavskaya: it turns 1 into 0 with probability α independently of all the other past and present events. Of course, if we start at the configuration "all zeros", the process remains there and if we start with "all ones", it immediately goes to "all zeros". However, computer simulations suggest that if α is small enough and we start with a chaotic configuration, where densities of zeros and ones are equal, the process remains chaotic, which suggests that the operator has a chaotic invariant distribution, which tends to a complete chaos when $\alpha \to 0$. By "complete chaos" we mean a product-measure, that is a distribution, in which all components are independent and every component is zero or one with equal probabilities. If $\alpha = 0$, the "complete chaos" really is invariant for this operator.

2.10. *Unsolved problem.* For any odd number $2k+1$ of arguments we can define a Boolean function *major* $(\cdot)$ of $2k+1$ arguments as follows: it equals 1 if at least $k+1$ of its arguments are ones and zero otherwise. Let us consider one-dimensional operator $R_\alpha^\beta\, D_{major}$ with $\Omega = \{0,1\}^{\mathbb{Z}}$, where D_{major} is defined by

$$(D_{major}\ x)_i = major\ (x_{i-k}, \ldots, x_{i+k}).$$

It follows from theorem 2.3 that this operator is ergodic as soon as

$$1 - \frac{1}{2k+1} \ < \ \alpha + \beta \ \leq \ 1.$$

Computer simulations and other considerations suggest that it is ergodic whenever $0 < \alpha, \beta < 1$, but it is not yet proved even for $k = 1$. The only positive result in this direction is a proof of an analogous statement for continuous time [Gray], but only for $k = 1$.

3. Eroders

Theorem 2.3 has provided us with plenty of ergodic cellular automata. Roughly speaking, a cellular automaton is ergodic as soon as interaction between its components is weak enough. What about non-ergodic cellular automata? We already have some examples of them: percolation operators with a small enough death rate. However, all of them are degenerate. The goal of this chapter is to present examples of non-degenerate non-ergodic cellular automata.

Although the space of our operators is discrete, we need to speak about continuous space now. A set in a linear space is called *convex* if with any two points it contains the segment with the ends at these points. Given a set in a plane, its *convex hull* is the intersection of all convex sets containing this set. The following figure gives two examples.

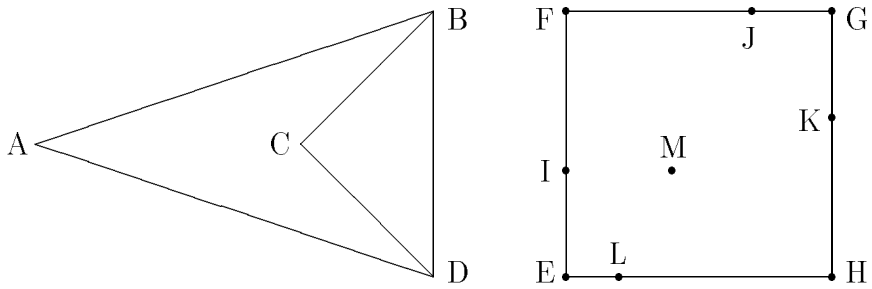

Figure 3.1. Examples of a set and its convex hull. $ABCD$ is a non-convex quadrangle. The triangle ABD is its convex hull. The quadrangle $EFGH$ is a convex hull of the finite set $\{E, F, G, H, I, J, K, L, M\}$.

We are mostly interested in convex hulls of finite sets. For example, convex hull of a point is the same point; convex hull of two points is the segment with the ends at these points; convex hull of three points is either the triangle with these points as vertices or the segment with two of these points as ends and the third point between them; convex hull of several points is either a segment (if all of these points belong to one line) or a convex polygon with some of these points as its vertices and other points inside or on the sides of it. If you put nails into a board at several points and turn a cord around them, it takes the form of the boundary of the convex hull of these points.

Let us call a configuration x *invariant* for a deterministic operator D if $Dx = x$. Let us call a configuration y a *finite deviation* from configuration x if the set of points v where $x_v \neq y_v$ is finite. Let us call D an x-eroder if x is invariant for D and for any finite deviation y from x there is t such that $D^t\, y = x$. In other words, D is an x-eroder if it "erodes" any finite distortion of x in a finite time. This notion is connected with the idea of stability. If D is an x-eroder, x is similar to a stable equilibrium like those which we observe in natural systems. There is another reason to study eroders: they are connected with phase transition as the next theorem shows.

Regretfully, the general problem of discerning eroders is algorithmically unsolvable, whence, if we want to obtain some positive results about them, we have to reduce our appetites. Let us restrict ourselves to the case $\Omega = \{0,1\}^U$, where $U = \mathbb{Z}^d$, and reduce our interest to x-eroders for two cases only: $x =$ "all zeros" and $x =$ "all ones"; we call them 0-eroders and 1-eroders. More than that, let us consider only uniform operators. This means that we choose a finite list of vectors $v_1, \ldots, v_n \in \mathbb{Z}^d$ and a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and define D as follows:

$$(D\, x)_v \;=\; f(x_{v+v_1}, \ldots, x_{v+v_n}) \quad \text{for all} \quad v \in \mathbb{Z}^d.$$

However, even after these restrictions the problem of discerning eroders remains unsolvable, even in the case $d = 1$. So we need an additional assumption. Let us assume that D is *monotonic*, which means that the function $f(\cdot)$ is monotonic, that is

$$x_1 \leq y_1, \ldots, x_n \leq y_n \implies f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n).$$

For this case we can present a criterion of eroders and existence of critical points, even in arbitrary dimension. For convenience let us assume that the function $f(\cdot)$ is not a constant. This excludes only two easy cases: $f(\cdot) \equiv 0$ and $f(\cdot) \equiv 1$. Let us denote $V = \{v_1, \ldots, v_n\}$ the set of neighhbor vectors. Let us call a subset S of V a *zero-set* if $(\forall i \in S : x_i = 0) \implies f(x_V) = 0$, where x_V means $x_{v_1}, \ldots, x_{v_n}$. We call a zero-set minimal if all its proper subsets are not zero-sets. Since V is finite, the set of its subsets is also finite, whence the set of minimal zero-sets is also finite and we may denote them $z_1, \ldots, z_n$. Since $f(\cdot)$ is monotonic and not a constant, it can be represented as

$$f(x_V) = \min_{i=1,\ldots,n} \max\left(\, x_v,\ v \in z_i \,\right).$$

We shall call this min-max representation. Notice that $f(x_V) = 0$ if and only if x_V has zeros at all points of at least one minimal zero-set: this is easy to prove in both directions. Now let us imbed

$\mathbb{Z}^d$ into a real space $\mathbb{R}^d$ and denote σ_0 the intersection of convex hulls of all minimal zero-sets.

Analogously we can do all the same for ones rather than zeros: call $S \subseteq V$ a *one-set* if $(\forall i \in S : x_i = 1) \Longrightarrow f(x_V) = 1$ and denote σ_1 the intersection of convex hulls of all minimal one-sets. In this case it makes sense to use max-min representation:

$$f(x_V) = \max_{i=1,\ldots,m} \min\Big(\, x_v,\ v \in o_i \,\Big),$$

where $o_1, \ldots, o_m$ is the list of minimal one-sets. Since these two cases are symmetric, we formulate our theorem only for the zero case. Let us denote $R^\beta = R_0^\beta$ the random operator, which turns any 0 into 1 with probability β independently, but does not turn ones into zeros.

Theorem 3.1 [T-80].
a) If σ_0 is empty, then D is a 0-eroder and there is a critical value β^* strictly between 0 and 1 such that the operator $R^\beta D$ is ergodic for $\beta > \beta^*$ and non-ergodic for $\beta < \beta^*$.
b) If σ_0 is not empty, then D is not a 0-eroder and $R^\beta D$ is ergodic for all positive β.

Before proving it, let us consider some examples. The first examples which come to mind are percolation operators defined by (6), all of which are 1-eroders as soon as $n \geq 2$. However, for reasons, which will become clear below, we are especially interested in operators, which are 0-eroders and 1-eroders at the same time. Here are two examples.

The NEC operator, where NEC stands for North-East-Center. In this case the deterministic operator D_{NEC} turns any $x \in \Omega$ into Dx defined as follows:

$$(Dx)_{i,j} = major\,(x_{i,j+1},\ x_{i+1,j},\ x_{i,j}) \quad \text{for all} \quad i, j, \qquad (10)$$

where the Boolean function $major\,(\cdot)$ of three arguments equals one if the majority of them are ones and zero if the majority of them are zeros. In other words,

$$major\,(a, b, c) = \begin{cases} 1 & \text{if the majority of } a, b, c \text{ are ones,} \\ 0 & \text{if the majority of } a, b, c \text{ are zeros.} \end{cases}$$

Notice that the function $major\,(\cdot)$ is monotonic. It can be written in the min-max form as follows:

$$f(x_V) =$$
$$\min\Big(\ \max(x_{0,1},\ x_{1,0}),\ \max(x_{0,1},\ x_{0,0}),\ \max(x_{1,0},\ x_{0,0})\,\Big).$$

For D_{NEC} there are three minimal zero-sets:

$$\{(0,1),(1,0)\}, \quad \{(0,1),(0,0)\}, \quad \{(1,0),(0,0)\}.$$

Their convex hulls are segments with these points as ends:

$$[(0,1),\ (1,0)], \quad [(0,1),\ (0,0)], \quad [(0,0),\ (1,0)].$$

The intersection of these three segments is empty. Thus the set σ_0 for D_{NEC} is empty and the operator should be a 0-eroder. You may prove it looking at the figure 3.5. In fact, it is a 1-eroder also. To prove this it is sufficient to notice that this function is 0-1 symmetric, which means that if all the three arguments change their values, the function also changes its value.

Another example is *flattening*, for which

$$(D_{flat}\ x)_{i,j} = \\ \min\ \Big(\ \max(x_{i,j},\ x_{i,j+1}),\ \max(x_{i+1,j}.\ x_{i+1,j+1})\Big). \qquad (11)$$

This formula already has min-max form. Its σ_0 and σ_1 are also empty (check!), so D_{flat} also is a 0-eroder and a 1-eroder. You can show it looking at the figures 3.3 and 3.4.

To present examples of non-eroders is still more easy; the dentity operator (which changes nothing) is neither 0-eroder nor 1-eroder. A more interesting example of a 0-noneroder (but 1-eroder) is given by formula (12).

Proof of theorem 3.1. The following scheme shows of which parts our proof consists. By a closed half-plane we mean one of two halves into which a plane is divided by a line, including this line. A zero-half-plane means a closed half-plane, which contains a zero-set.

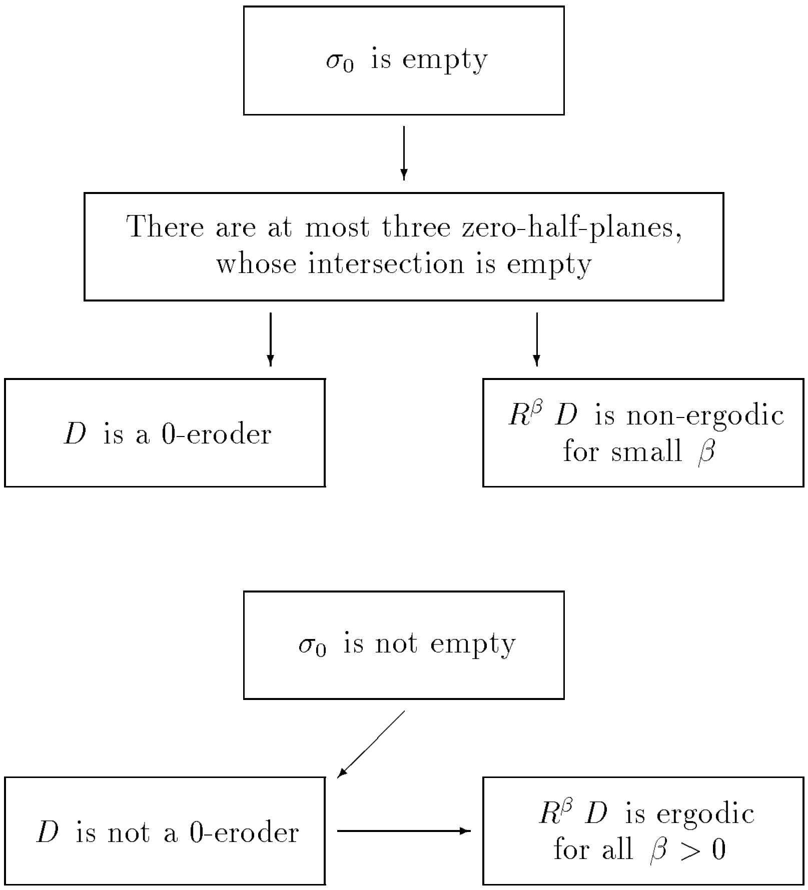

Figure 3.2. Scheme of proof of theorem 3.1.

Every arrow in this scheme represents an argument. We shall call these arguments propositions. But first of all we need the following theorem.

Theorem 3.2. (A version of Helly's theorem.) If there is a finite family of convex sets in $\mathbb{R}^d$ such that every $d+1$ of them have a common point, then all the sets in the family have a common point.

We don't need this theorem in general, it is sufficient to prove it for the case when all these convex sets are closed half-planes. We leave to the reader to prove the following special case of Helly's theorem for $d=1$: if there are several closed half-lines in a line, every two of which have a non-empty intersection, then all of them have a non-empty intersection. (A closed half-line is one of the two halves into which a line is cut by a point, including this point.) Also we need the following statement: if there are two closed convex sets in a plane, which do not intersect, then there is a line in this plane, which separates them, so that these sets are on different sides of this line. For our case, when all the sets in question are intersections of several closed half-planes, this statement is evident and we also leave its proof to the reader.

Now let us prove by contradiction that if there is a finite family of closed half-planes in a plane such that every three of them have a common point, then all these closed half-planes have a common point. Let n be the smallest number of closed half-planes in a counter-example, and let $C_1, \dots, C_n$ be closed half-planes in a plane whose intersection is empty although every three of them have a non-empty intersection. Since n is minimal with this property, the intersection $I = C_1 \cap \dots \cap C_{n-1}$ is non-empty. However, I has no common points with C_n. Then there is a straight line x in the plane, which separates them, so that I and C_n are on different sides of it. Now for all $i = 1, \dots, n-1$ we denote D_i the intersection of C_i with this line x. Let us prove that every two of the sets $D_1, \dots, D_{n-1}$ have a common point. Let us take some sets C_i and C_j, We know that C_i, C_j and C_n have a common point. Therefore the intersection $C_i \cap C_j$ has a point on one side of x (where I is) and a point on the other side (where C_n is). Hence, since $C_i \cap C_j$ is convex, it has a common point with x and this point belongs to $D_i \cap D_j$, which therefore is non-empty. Now we can apply Helly's theorem for the one-dimensional case to the sets $D_1, \dots, D_{n-1}$ because all of them are closed half-lines (unless they are empty or coincide with this line, which is easy to handle). Since we have proved that every two of them have a common point, all of them must have a common point. Let us call this point y. But every D_i is a subset of C_i, whence y belongs to all $C_1, \dots, C_{n-1}$, whence it belongs to their intersection I. So the line x has a common point y with I, which contradicts our choice of the line x. This contradiction proved Helly's theorem in the case in which we need it.

Now we start to prove theorem 3.1.

Proposition 3.1. If σ_0 is empty, then there are at most three zero-half-planes, whose intersection is empty.

Proof. Every convex hull of a zero-set can be represented as an intersection of several closed half-planes, all of which are zero-half-planes. Therefore, σ_0 can be represented as an intersection of several zero-half-planes. Since it is empty, from Helly's theorem at most three of them have an empty intersection. *Proposition 3.1 is proved.*

Proposition 3.2. If there are at most three zero-half-planes, whose intersection is empty, then D is a 0-eroder.

What is important here is the minimal number of zero-half-planes, whose intersection is empty. On the plane this number is either 2 or 3 and these cases should be considered separately. To hit at the idea in each case it is sufficient to examine our two examples: flattening for the former case and D_{NEC} for the latter case.

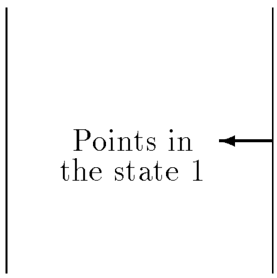

Figure 3.3. Why flattening is a 0-eroder. The set of points in the state 1 is between these lines until it diasppears. The left line does not move. The right line moves in the direction of arrow as time goes on.

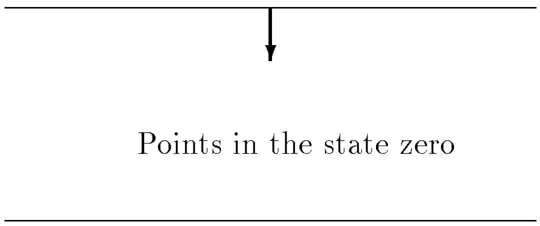

Figure 3.4. Why flattening is a 1-eroder. The set of points in the state 0 is between these lines until it disappears. The lower line does not move. The upper line moves in the direction of arrow as time goes om.

For flattening there are two zero-half-planes whose intersection is empty:

$$\{(i,j) : i \leq 0\} \quad \text{and} \quad \{(i,j) : i \geq 1\}.$$

Based on this, let us show that flattening is a 0-eroder. Given a configuration x with a finite $I_1(x)$, let us draw two vertical lines such that this set is between them. Then $I_1(D^t\, x)$ is also between two vertical lines: one of these lines remains the same and the other line moves towards it as t grows. Thus at every time step the distance between these lines decreases and the configuration becomes "all zeros" in a time equal to the initial distance between the lines.

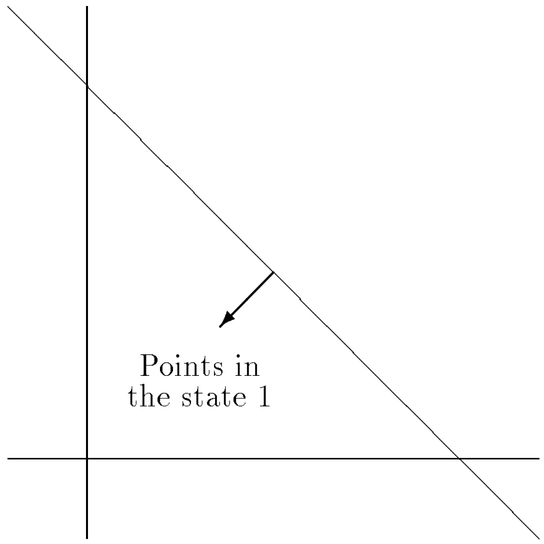

Figure 3.5. Why D_{NEC} is a 0-eroder. The set of points in the state 1 is in the triangle formed by these three lines until it disappears. The horizontal and vertical lines do not move. The inclined line moves in the direction of arrow as time goes on. Due to its 0-1 symmetry, D_{NEC} is a 1-eroder for the same reason.

For D_{NEC} there are three zero-half-planes, whose intersection is empty:

$$\{(i,j) : i \leq 0\}, \quad \{(i,j) : j \leq 0\} \quad \text{and} \quad \{(i,j) : i+j \geq 1\}.$$

Based on this, let us show that D_{NEC} is a 0-eroder. For any configuration x with a finite set $I_1(x)$ we draw three lines so as to place this set into a triangle surrounded by these lines: one horizontal, one vertical and one with the slope -1. After every application of D_{NEC} the set $I_1(D_{NEC}^t\ x)$ remains between the lines, the horizontal and vertical lines remaining the same, but the line with the slope -1 moving towards their intersection. So the set $I_1(D_{NEC}^t\ x)$ shrinks to an empty set and the configuration becomes "all zeros" in a time proportional to its initial size. Using these ideas you can prove proposition 3.2.

Proposition 3.3. If there are at most three zero-half-planes, whose intersection is empty, then $R^\beta\ D$ is non-ergodic for small enough $\beta > 0$.

A general proof needs some sophisticated technique: either branching analogs of contours [T-80], or renormalization [Bra+Gra]. [Leb+Mae+Spe] contains a good explanation of the branching method for the NEC operator. We shall prove proposition 3.3 only for one particular case: the flattening operator $P = R^\beta\, D_{flat}$, where D_{flat} was defined by (11). Notice that δ_1 is invariant for P, so it is sufficient to prove that $P^t\,\delta_0$ does not tend to δ_1 when $t \to \infty$ for β small enough. In fact we shall estimate from above the density of zeros in measures $P^t\,\delta_0$ uniformly in t.

Let us use triple $(i,\ j,\ t)$ when we speak of the point $(i,\ j)$ at time t. We shall imagine this triple as a point in a three-dimensional integer space where the axes i, j are horizontal and the axis t goes upward. We may also denote a point $(i,\ j,\ t)$ by one letter, say A, and in this case we write $i = i(A),\ j = j(A),\ t = t(A)$. Let $x(i,\ j,\ t)$ equal 1 if there is a particle at $(i,\ j,\ t)$ and equal 0 otherwise. Also let us use independent random variables $y(i,\ j,\ t)$, which equal 1 if the random operator turned zero into one at this point and equal 0 otherwise. Every $y(i,\ j,\ t)$ equals one with probability β and zero with probability $1-\beta$ independently of other variables $y(\cdot)$. Thus we can define the variables $x(i,\ j,\ t)$ in the following inductive way:

Base of induction: $x(i,\ j,\ 0) = 0$ for all i, j.

Induction step: $x(i,\ j,\ t) = \max(y(i,\ j,\ t),\ z(i,\ j,\ t))$, where the intermediate variable $z(i,\ j,\ t)$ is (i,j)-th component of the result of application of D_{flat} to the configuration at time $t-1$, that is

$$
\begin{aligned}
& z(i,j,t) \;= \\
& \min(\max(x(i\quad ,j,\ t-1),\ x(i\ \ ,j+1,\ t-1)), \\
& \qquad \max(x(i+1,j,\ t-1).\ x(i+1,j+1,\ t-1))).
\end{aligned}
$$

This pseudo-code expresses the same idea:

```
for all (i,j) ∈ ℤ² do
    x(i, j, 0) ← 0
for t = 1 to ∞ do
    for all (i,j) ∈ ℤ² do
        x(i,j,t) ← (D_flat x(t − 1))(i,j)
        if rnd < β then x(i,j,t) ← 1
```

Here $(D_{flat}\ x(t-1))(i,j)$ is the (i,j)-th component of the result of application of D_{flat} to the configuration $x(t-1)$, whose components are $x(i,j,t-1),\ i,j \in \mathbb{Z}$.

Let us estimate from above the probability that there is a particle at $(0,\ 0)$ at time T, that is $x(0,\ 0,\ T)=1$. We shall cover this event by several events and estimate the sum of their probabilities. To every one of these events there will correspond a special path. If k-th and $(k+1)$-th vertices of this path are $(i_k,\ j_k,\ t_k)$ and $(i_{k+1},\ j_{k+1},\ t_{k+1})$, then we denote $\Delta_k=(\Delta i_k,\ \Delta j_k,\ \Delta t_k)$, where

$$\Delta i_k=i_{k+1}-i_k,\quad \Delta j_k=j_{k+1}-j_k,\quad \Delta t_k=t_{k+1}-t_k.$$

Let us take an arbitrary realization of our process and construct the event to which it belongs along with the corresponding path. We shall proceed inductively, at every step constructing some event and some path and proving the following induction assumptions about them:

a) This path leads from $(0,\ 0,\ T)$ to $(1,\ 0,\ T)$.

b) This path has steps only of three following types:

$$\begin{cases}\textbf{down}\text{, having } \Delta i=0 \text{ and } \Delta t=-1\,.\\ \textbf{horizontal}\text{, having } \Delta i=1 \text{ and } \Delta t=0\,.\\ \textbf{up}\text{, having } \Delta i=-1 \text{ and } \Delta t=1\,.\end{cases}$$

In all the three cases $|\Delta j|\le 1$.

c) If a horizontal step starts at $(i,\ j,\ t)$, then $x(i,\ j,\ t)=1$.

Base of induction. The event is "$x(0,\ 0,\ T)=1$" and the path is $(0,\ 0,\ T)\to(1,\ 0,\ T)$. It is evident that all the assumptions are fulfilled.

When we stop: when our path has the following property:

d) for every vertex $(i,\ j,\ t)$, where a horizontal step starts, $y(i,\ j,\ t)=1$.

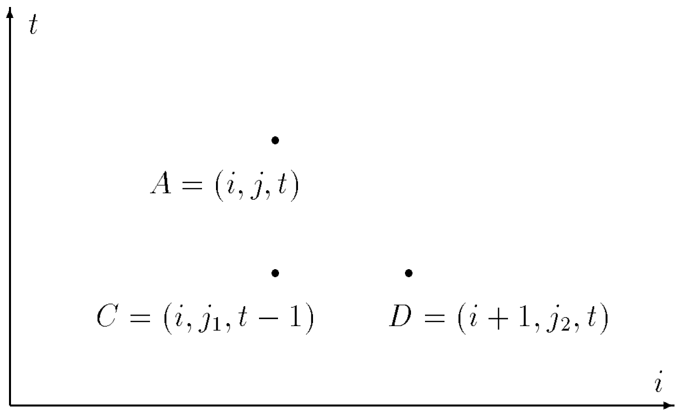

Figure 3.6. An illustration of one induction step. The axis j is not drawn, it is perpendicular to the paper. If $x(i,\ j,\ t) = 1$, but $y(i,\ j,\ t) = 0$, this value of x must be "inherited" from the previous time step: there must be points $(i, j_1, t-1)$ and $(i+1,\ j_2,\ t-1)$, where x equals one, j_1 and j_2 being equal either j or $j+1$.

Induction step. Suppose that there is a vertex $A = (i,\ j,\ t)$, where a horizontal step AB starts, but $y(i,\ j,\ t) = 0$. By induction assumption, $x(i,\ j,\ t) = 1$. Since the variable y at this point is zero, the value of x at this point is inherited from the previous time step. Looking at the figures 3.3 and 3.6 will help you to realize that there must be two other points C and D such that

$$t(C) = t(D) = t(A) - 1, \quad i(C) = i(A) \quad \text{and} \quad i(D) = i(A) + 1.$$

Then we define another vertex E by the rule: the vector from D to E is the same as the vector from A to B. After that we change our path as follows: instead of going straight from A to B, we go from A to C, then to D, then to E and thence to B. In other words, we insert $C,\ D,\ E$ between A and B into the sequence of vertices of our path. It is easy to prove that the new path also satisfies the induction assumptions a), b), c). This induction process cannot last forever because $x(i, j, 0) \equiv 0$. When it stops, we have a path satisfying the requirements a), b), c) and d). but this path may not be self-avoiding yet. We want a self-avoiding path with all these properties. To obtain it, we use "delooping" similar to that which we used when proving lemma 1.1. If the path which we have is not yet self-avoiding, it visits some vertex twice and makes a loop between these visits. We eliminate this loop (including only one of these visits) from our path, thus obtaining a shorter path, which has all the properties a), b), c) and d). So we do until we get a self-avoiding path with these properties.

Our event is that $y(i,\ j,\ t) = 1$ for all those vertices of our path where horizontal steps start. Since the path is self-avoiding, probability of this event is not greater than β^k, where k is the number of horizontal steps in the path. Notice that the number of **down** steps is equal to the number of **up** steps and both equal $k-1$. So the length of the path is $3k-2$. The number of possible paths with length $3k-2$ does not exceed C^{3k-3} where C is the number of different vectors $(\Delta i,\ \Delta j,\ \Delta t)$ allowed by condition b). Thus the sum of probabilities of all events does not exceed

$$\sum_{k=1}^{\infty} C^{3k-3}\,\beta^k \;=\; \frac{\beta}{1-C^3\beta}.$$

For β small enough this sum is less than 1, whence measures $P^t\,\delta_0$ do not tend to δ_1. *Proposition 3.3 is proved.*

Proposition 3.4. If σ_0 is not empty, then D is not a 0-eroder.

We shall prove this for the case when σ_0 contains the origin. In this case we shall present a configuration x, which is a finite deviation from "all zeros", such that $I_1(D\,x) \supseteq I_1(x)$. From monotonicity this implies that $I_1(D^t\,x)$ is non-empty for all t. Any set $S+v=\{i+v,\ i\in S\}$, where v is a vector, is called a *shift* of S. Let us call a set S *obtuse* for another set Q if any shift of S, which intersects $\mathrm{conv}\,(Q)$, intersects Q also. The following figure illustrates this.

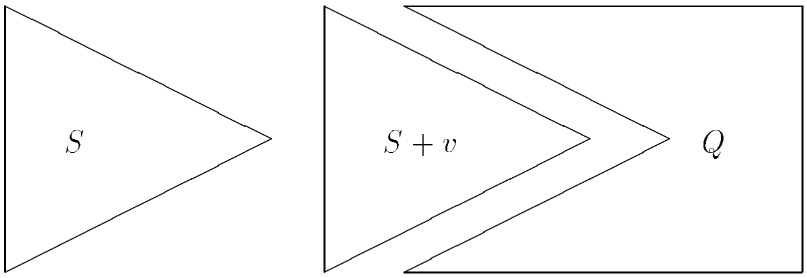

Figure 3.7. The set S is not obtuse for the set Q because its shift $S+v$ intersects the convex hull of Q without intersecting Q.

Theorem 3.3. Carathéodory's theorem for bounded 2-dimensional sets. Let S be any set in a plane. Then a point p belongs to $\mathrm{conv}\,(S)$ if and only if it belongs to $\mathrm{conv}\,(\{p_1,\ p_2,\ p_3\})$, where $p_1,\ p_2,\ p_3$ are some points in S (some of which may coincide). In other words,

$$\mathrm{conv}\,(S) = \bigcup_{p_1,\ p_2,\ p_3\,\in S} \mathrm{conv}\,(\{p_1,\ p_2,\ p_3\}).$$

We leave its proof to the reader.

Lemma 3.4. For any set S in a plane the set $-3\,\mathrm{conv}\,(S)$ is obtuse for S.

We leave to the reader proof of this lemma when S consists of one, two or three points and, assuming that for these cases the lemma is already proved, prove it in general. Let us take any shift $S+v$ of S, assume that it does not intersect $-3\,\mathrm{conv}\,(S)$ and prove that $\mathrm{conv}\,(S)+v$ also does not intersect $-3\,\mathrm{conv}\,(S)$. Let us take any point $p\in\ \mathrm{conv}\,(S)$. Now, due to theorem 3.3, there are $p_1,\ p_2,\ p_3\in S$ such that $p\in\ \mathrm{conv}\,(\{p_1,\ p_2,\ p_3\})$. By assumption, $S+v$ does not intersect $-3\,\mathrm{conv}\,(S)$, whence $p_1+v,\ p_2+v,\ p_3+v$ do not belong to $-3\,\mathrm{conv}\,(S)$. Since our lemma is already proved for sets containing at most three points, we can conclude that $\mathrm{conv}\,(\{p_1+v,\ p_2+v,\ p_3+v\})$ does not intersect $-3\,\mathrm{conv}\,(S)$, whence p does not belong to $-3\,\mathrm{conv}\,(S)$. *Lemma 3.4 is proved.*

Now notice that for any $S_1,\ S_2,\ z\subseteq\mathbb{R}^2$: If S_1 is obtuse for z and S_2 is non-empty, then S_1+S_2 is obtuse for z, where $+$ means vector sum:

$$S_1+S_2=\{s_1+s_2:s_1\in S_1,\ s_2\in S_2\}.$$

The number of minimal zero-sets is finite since all of them are subsets of V. So we can denote them $z_1,\ldots,z_n$. For every minimal zero-set z_i we have a bounded set S_i obtuse for it. Then their vector sum $S_1+\cdots+S_n$ is also bounded and obtuse for all minimal zero-sets. Let us add a large enough sphere S_0 to be sure that the intersection of the total vector sum S with $\mathbb{Z}^2$ is non-empty and define

$$S=S_0+S_1+\cdots+S_n.$$

Thus we have a bounded set S, which is obtuse for all zero-sets. Let us prove that the configuration x, whose $I_1(x)=S\cap\mathbb{Z}^2$, is not eroded by D. In fact we shall prove that $I_1(D\ x)\supseteq I_1(x)$. Assume the opposite: there is a point v such that $x_v=1$, but $(D\,x)_v=0$. Since $(D\,x)_v=0$, there must be a minimal zero-set z such that $x_{v+i}=0$ for all $i\in z$. Therefore $z+v$ does not intersect S. Since S is obtuse for z, the convex hull of $z+v$ also does not intersect S. Since σ_0 contains the origin, the convex hull of z also contains the origin, so the convex hull of $z+v$ contains v, whence v does not belong to S, which contradicts our assumption that $x_v=1$. *Proposition 3.4 is proved.*

Proposition 3.5. If D is not a 0-eroder, then $R^\beta D$ is ergodic for all positive β.

For this purpose we estimate the probability that the 0-th component is zero after t applications of $R^\beta D$ to any initial configuration and prove that this probability tends to zero when $t \to \infty$. Since D is not a 0-eroder, there is a finite deviation x from "all zeros". not eroded by it. So $I_1(D^t x)$ is not empty for all natural t and we can choose a point p_t in it. For every $u \in [1, t]$ let us consider the event:

"At the time $u \in [1,t]$ the random operator R^β turns all components of $I_1(x_1) - p_{t-u}$ into ones."

This event is sufficient for 1 to be at the point 0 at time t. Only if none of these events took place, 0 may be at the point 0 at time t. But these events are independent from each other (because they pertain to different times) and the probability of each is β^C where C is the number of elements in $I_1(x)$. Therefore for every of these events the probability to happen is β^C and the probability not to happen is $1-\beta^C$. So the probability that none of these events happens is

$$\left(1-\beta^C\right)^t,$$

which tends to zero when t tends to ∞. Thus the probability to have zero at the origin tends to zero as $t \to \infty$. The same is true for any point and this is sufficient for zeros to die out.

Now let us remember the promise we gave in the beginning of this chapter: to present non-degenerate non-ergodic cellular automata.

Theorem 3.4. If D is a monotonic deterministic operator defined by (9) and it is a 0-eroder and a 1-eroder, then $R_\alpha^\beta D$ has at least two different invariant measures for small enough α and β.

We shall prove this theorem in the last chapter. Let us mention here results of computer modelling of one of those operators, about which this theorem speaks, $R_\alpha^\beta D_{NEC}$. It was first studied by computer simulation in [Pet+Pia+Vas] in the symmetric case $\alpha = \beta \le 1/2$ and phase transition was observed:[5] when $\alpha = \beta$ was close to $1/2$, the operator was ergodic, that is tended to one and the same regime from all initial conditions, but for $\alpha = \beta$

[5] Due to monotonicity, it was sufficient to consider only two initial conditions "all zeros" and "all ones", and due to symmetry it was sufficient to consider only one of them.

small enough the operator was non-ergodic, i.e. it "remembered" the initial condition: if simulation started from "all zeros", zeros prevailed all the time, if it started from "all ones", ones prevailed all the time. Since this operator has two parameters α and β, instead of a critical point there is a critical curve which was studied experimentally in [Ben+Gri]. Let us use parameters

$$\text{amplitude} = \beta + \alpha, \quad \text{bias} = \frac{\beta - \alpha}{\beta + \alpha}.$$

Figure 3.7 represents ergodic and non-ergodic regions as observed by computer simulation [Ben+Gri].

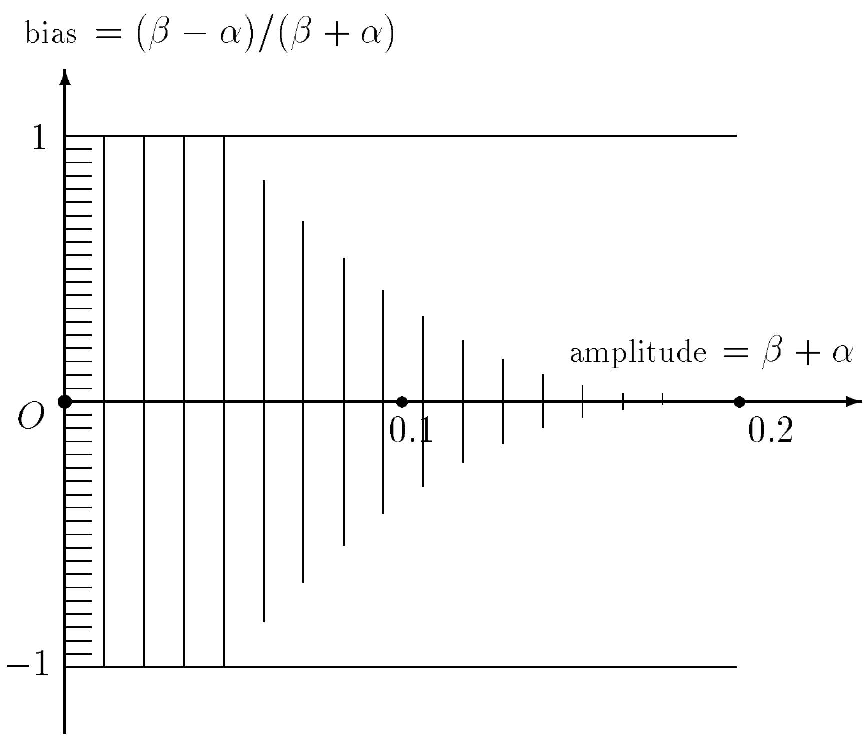

Figure 3.7. Schematic representation of the zones of ergodicity and non-ergodicity for the NEC operator according to computer simulation [Ben+Gri]. *The empty area shows where ergodicity was observed in the experiment. The vertically shaded area shows where non-ergodicity was observed. The horizontally shaded area shows where non-ergodicity is proved. Ergodicity is proved for amplitude close to* 1, *beyond the right border of this figure.*

Notes.

3.1. *Exercise.* Pay attention that σ_0 may contain no integer points and still be non-empty. Let $D : \Omega \to \Omega$, where $\Omega = \{0,1\}^U$,

$U = \mathbb{Z}^2$, be defined as follows:

$$(Dx)_{i,j} = \min(\max(x_{i,j},\ x_{i+1,j+1}),\ \max(x_{i+1,j},\ x_{i,j+1})).$$

a) What is σ_0 in this case?
b) Let us denote x the configuration such that

$$I_1(x) = \{(0,0),\ (1,0),\ (0,1),\ (1,1)\}.$$

What are sets $I_1(D^t\, x)$ for t=1,2,3,...?

3.2. *Exercise.* Prove another version of Helly's theorem: If there is a family of bounded convex sets in $\mathbb{R}^d$ such that every $d+1$ of them have a common point, then all the sets in the family have a common point. (The family does not need to be finite.) You can find a proof of a very general version of Helly's theorem and many related facts in [Rockafellar].

3.3. *Exercise.* Prove that the following statement is false: "If there is a family of convex sets in $\mathbb{R}^d$ such that every $d+1$ of them have a common point, then all the sets in the family have a common point."

3.4. Let us consider uniform monotonic deterministic operators acting on $\Omega = \{0,1,\ldots,m\}^U$. where $U = \mathbb{Z}^d$. We call an operator a 0-eroder if it "erodes" all finite deviations from "all zeros" by turning them into "all zeros".
Solved problem. [Galperin] contains a criterion of 0-eroders for arbitrary finite set of states, but only for dimension $d = 1$.
Unsolved problem: To obtain a similar result for dimension 2 and three states at every point.

3.5. *Solved problem.* Theorem 3.1 states a connection between eroders and existence of critical values. This connection is not true for greater numbers of states of components. This is shown by the following one-dimensional counter-example, where every point has three states: 0,1 and 2. In this case $\Omega = \{0,1,2\}^{\mathbb{Z}}$ and uniform deterministic operator D is defined by the rule

$$\forall x \in \Omega,\ v \in \mathbb{Z}\ :\ (Dx)_v = f(x_{v-1},\ x_v,\ x_{v+1}),$$

the function $f(\cdot)$ being defined as follows:

$$f(x,\ x_0,\ x_1) = \begin{cases} 1 & \text{if } x_{-1} = 1,\ x_0 = x_1 = 2, \\ 0 & \text{if } x_{-1} = x_0 = 1,\ x_1 = 0, \\ (x_{-1} + x_0 + x_1)/3 & \text{rounded to the} \\ & \text{nearest integer number} \\ & \text{in all the other cases.} \end{cases}$$

a) Check that the function $f(\cdot)$ is monotonic.
b) Let us call a configuration x a finite deviation from "all zeros" if the set $\{v \in \mathbb{Z} \ : \ x_v \neq 0\}$ is finite. Prove that D is an eroder that is for any finite deviation x from "all zeros" there is t such that $D^t x =$ "all zeros".
c) Let R^β be random operator, which turns any component into the state 2 with probability β independently of others. Prove that for any $\beta > 0$ the operator $R^\beta D$ is ergodic [Discr].

3.6. *Unsolved problem.* Let us consider the finite-space version of NEC operator, that is let $U = \mathbb{Z}_m^2$. Let us consider its invariant measure, which is unique as soon as $0 < \alpha, \beta < 1$. For any $n \in [0, m^2]$ let us denote $P(n)$ the probability (according to this invariant measure) that there are n ones in the configuration. Prove that for large values of m the shape of distribution of $P(n)$ is qualitatively different for different values of α and β, namely:
a) If $\alpha + \beta$ is close to 1, the distribution of $p(n)$ is unimodal: it is largest when n is close to $m^2/2$.
b) If α and β are small, the distribution of $p(n)$ is bimodal: it is largest when n is close to pm^2 and $(1-q)m^2$ where p and q tend to zero when α and β tend to zero. The same is probably true for all monotonic uniform operators which are 0-eroders and 1-eroders at the same time.

3.7. *Unsolved problem.* Along with the NEC operator, the article [Pet+Pia+Vas] wrote about another operator $R_\alpha^\beta \, D_{major}$, acting on the same space, where D_{major} is defined by

$$(D_{major}\ x)_{(i,j)} = major(x_{i,j},\ x_{i+1,j},\ x_{i,j+1},\ x_{i-1,j},\ x_{i,j-1}), \quad (12)$$

where the Boolean function $major(\cdot)$ of five arguments equals 0 if the majority of its arguments equals 0 and equals 1 if the majority of its arguments equals 1. Since D_{major} is neither 0-eroder nor 1-eroder, theorem 3.4 cannot be used. However, computer modelling suggested that this operator is non-ergodic for small enough $\alpha = \beta$. Also it seems plausible that this operator is non-ergodic whenever $\alpha \neq \beta$. Both statements are unproved.

4. Ergodicity problem for cellular automata is unsolvable

This chapter is devoted to algorithmic problems related to cellular automata. Let us think, what do we really want to achieve in dealing with cellular automata, in particular with the study of their ergodicity. Mathematics is an abstract science and we, mathematicians, want to prove general theorems. In the present case we want to have criteria of ergodicity for as large classes of cellular automata as possible. However, it is known that some general problems in all areas of mathematics *cannot* be solved in the algorithmical sense. It is only natural that in dealing with cellular automata we face such situations very often, because our object is very general. When we meet an undecidable problem, it means that we are working close to the boundaries of natural possibility. This moves us to treat with more respect those partial results which we have obtained: perhaps, they are close to what can be done at all.

In this chapter we shall show that the problem of deciding, which cellular automata are ergodic and which are not, is algorithmically unsolvable for a certain class of them. We shall use one of formalizations of algorithms, namely Turing machines, because they are most similar to cellular automata. In fact we shall use the following class of Turing machines with one head and one bi-infinite tape. A Turing machine of this class consists of a head and a tape. The tape is infinite in both directions and consists of cells enumerated by integer numbers. Every cell can be in several states. The set G of states is one and the same for all cells. The head also has a finite set $H \cup \{stop\}$ of states, where one state, called *stop* , plays a special role described below. At every step of the discrete time the head observes one cell of the tape as shown in figure 4.1.

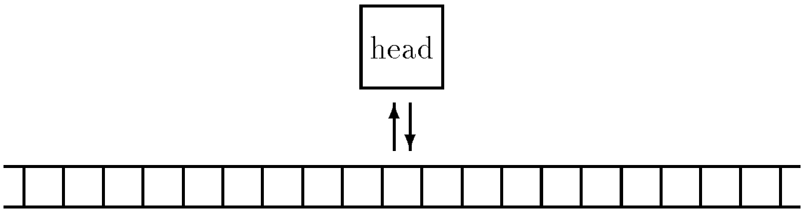

Figure 4.1. Turing machine. At every step of the discrete time the head observes one cell, exchanges information with it and then moves to another cell.

Also we choose three functions:

$$F_{\text{tape}} \ : \ G \times H \to G,$$

$$F_{\text{head}} \;:\; G \times H \to H \cup \{stop\},$$
$$F_{\text{move}} \;:\; G \times H \to \{\leftarrow, \rightarrow\}.$$

When the machine starts, the tape is "empty", which means that all cells are filled with the initial symbol $g_1 \in G$. The head is in the initial state $h_1 \in H$ and the head observes the 0-th cell of the tape. At every step the head simultaneously writes into that cell of the tape, which it observes, a new symbol according to the function F_{tape}, goes to a new state according to the function F_{head}, and moves one cell left or one cell right along the tape according to the values $\leftarrow, \rightarrow$ of the function F_{move} respectively, the arguments of all the three functions being the symbol in the presently observed cell of the tape and the present state of the head. The machine stops when and if the head reaches the state *stop* . (That is why we don't need to define our functions when the head is in the state *stop* .) It is well-known that the problem of deciding, which of these machines ever stop, is algorithmically unsolvable, that is there is no algorithm capable to predict for all Turing machines, which of them ever stop having started with empty tape. This famous theorem is described in many books including [Minsky, Wang]. We shall use it to prove another theorem about undecidability - about cellular automata.

Let us define a certain class of cellular automata. In the most part of mathematics a theoretical result is the better the more general it is, that is the larger is the class of objects under consideration. However, when we prove algorithmic unsolvability, it is the other way round: the result is the more valuable the smaller is the class of objects for which it is proved. For this reason, we minimize our class as much as possible: the arbitrary number n of states of a single component is the only source of infinitness of our class, everything else is minimized: the dimension is 1, the interaction is only between nearest neighbors and all the transition probabilities are either 0 or 1/2 or 1.

Our configuration space is $\{1, \ldots, n\}^{\mathbb{Z}}$, where n is a natural parameter. A cellular automaton is an operator, which is a superposition of two operators: first deterministic D, then random R. Our deterministic operator is determined by a function $f : \{1, \ldots, n\}^3 \to \{1, \ldots, n\}$ in the following way: it transforms any configuration x into $D\,x$, where

$$(D\,x)_i \;=\; f(x_{i-1},\; x_i,\; x_{i+1}) \quad \text{for all} \quad i \in \mathbb{Z}. \tag{13}$$

Our random operator R is very simple: all components of a configuration, which are not in the state n, turn into the state 1 with probability $1/2$ independently of each other. An operator P is called *ergodic* if the dsitribution $P^t\mu$ tends to one and the same limit dsitribution from all initial conditions μ.

Theorem 4.1. There is no algorithm to decide which cellular automata described above are ergodic and which are not [Kurdyumov, Discr].

Our method of proof consists in the following: for every Turing machine of the class described above we construct a cellular automaton of the class described above, which is ergodic if and only if that Turing machine stops having started on an empty tape. This is sufficient to prove that the problem of deciding which of our cellular automata are ergodic is unsolvable because if it were solvable, the problem of deciding which of Turing machines stop would be solvable also, but it is known that it is not. In more detail: if the ergodicity problem were solvable, then we could take a Turing machine, construct the corresponding cellular automaton, apply to it that hypothetical deciding procedure, decide whether it is ergodic or not, and conclude whether the Turing machine stops or not.

Thus for every Turing machine T we shall construct an operator P which is ergodic if and only if T stops. In fact, P imitates the functioning of T in the following way: under its action, all components at any time randomly (with probability $1/2$) turn into heads of T in initial condition. Every head marks its territory with brackets and imitates the functioning of T on its territory. This functioning may be interrupted by other heads, but measures are taken to prevent the heads from collaborating: as soon as a head gets any sign of presence of another head, it commits a suicide. If M never stops, the process remains in this messy regime forever. If M stops, some components go to a special *final* state, which starts an "epidemy" by turning their neighbors into the same state, so that the process tends to the $\delta(\text{all } final)$, measure concentrated in the configuration "all components are in the *final* state". In fact, this measure is invariant in all cases, but the process tends to it from all initial conditions only if M stops.

Thus, for any Turing machine M of the class described above we shall construct an operator P, which is ergodic if and only if M stops. We set

$$S = S_{left} \times S_{right} \times S_{tape} \times S_{head},$$

where

$$S_{left} = S_{right} = \{0, 1\}, \qquad S_{tape} = G, \qquad S_{head} = H \cup \{0,\ stop\}.$$

Accordingly, we write a generic element of S as

$$x = (left(x),\ right(x),\ tape(x),\ head(x)).$$

We say that a state x has a *left bracket* if $left(x) = 1$ and that it has a *right bracket* if $right(x) = 1$.

We call x a *no-head* if $head(x) = 0$ and a *head* otherwise. We call x a *stop-head* if $head(x) = stop$.

The state $(0,\ 0,\ g_1,\ 0)$ is called *empty*, the state $(1,\ 1,\ g_1,\ h_1)$ is called *newborn* and the state $(0,\ 0,\ g_1,\ stop)$ is called *final*.

For brevity we shall write $F_*(x) = F_*(tape(x),\ head(x))$, where $*$ means 'tape', 'head' or 'move'.

We say that a head x wants to move left or to move right when $F_{\text{move}}(x)$ equals $\leftarrow$ or $\rightarrow$ respectively.

We need all these states to make our process imitate the functioning of Turing machine M. The tape component imitates what is written on the tape, the head component imitated what is in the head or its absence if it is zero. The left and right brackets are necessary to exclude interference of the heads.

Our operator P is a superposition $P = RD$, where R turns any state except final into the newborn state with a probability $1/2$ independently. It remains to define the deterministic operator D, that is to define the function $f(\cdot)$ in formula (13). Its definition consists of several rules.

Rule 0. If x or y or z is a stop-head, then $f(x, y, z) = final$.

Formulating all the other rules, we assume that neither x nor y nor z is a stop-head. We call a triple $(x, y, z) \in S^3$ *normal* if at most one of x, y, z is a head.

Rule 1. Whenever the triple (x, y, z) is not normal, $f(x, y, z) = empty$.

In all the following rules we assume that the triple (x, y, z) is normal.

Rule 2. If all of x, y, z are no-heads, then $f(x, y, z) = y$.

All the subsequent rules form three groups depending on which of the three arguments is a head: center, denoted y, or its left neighbor, denoted x, or its right neighbor, denoted z. The center-rules:

Rule 1-center. If y is a head which wants to move left, then

$$f(x,y,z) = (0,\ right(y),\ F_{\text{tape}}(y),\ 0).$$

Rule 2-center. If y is a head which wants to move right, then

$$f(x,y,z) = (left(y),\ 0,\ F_{\text{tape}}(y),\ 0).$$

Since the left rules and the right rules are symmetric, we shall omit the right ones. The left-rules:

Rule 1-left. If x is a head, which wants to move right and has a right bracket, then

$$f(x,y,z) = (0,\ 1,\ g_1,\ F_{\text{head}}(x)).$$

Rule 2-left. If x is a head, which wants to move right and has no right bracket and y has no left bracket, then

$$f(x,y,z) = (0,\ right(y),\ tape(y),\ F_{\text{head}}(x)).$$

Rule 3-left. If x is a head, which wants to move right and has no right bracket, but y has a left bracket, then $f(x,y,z) = y$.

Rule 4-left. If x is a head, which wants to move left, then $f(x,y,z) = y$.

The right-rules are analogous to left-rules, only with right and left permuted.

Our operator D is defined. To make the operator R satisfy the promised condition, it is sufficient to choose n equal to the cardinality of S and enumerate the states of S so that the newborn state gets number 1 and the final state gets number n.

Lemma 4.1. Operator $P = RD$ is ergodic if and only if the Turing machine M stops.

Proof of lemma 4.1. Due to the *rule 0*, the measure $\delta(\text{all } \mathit{final})$ concentrated in the configuration "all components are in the *final* state" is invariant for P. Therefore P is ergodic if and only if our process tends to δ(all *final*) from any initial configuration. Now let us argue in two directions.

One direction: Let us suppose that M stops after T steps and prove that our process tends to δ(all *final*) from any initial configuration. Let us consider a region $[s_0-2T,\ s_0+2T] \subset \mathbb{Z}$, where s_0 is any integer number. If a stop-head is present there, it turns into $final$, which expands in both directions due to *rule 0.* If there is no stop-head there, then the following scenario has a positive probability: First, at some time t_0 births occur in all sites in the range $[s_0-2T,\ s_0+2T]$. At the next time step all of these sites become empty. At the next time step birth occurs at the middle site s_0 and this is the only birth that occurs in the space-time region

$$\{(s,t) \mid s_0-2T+(t-t_0) \le s \le s_0+2T-(t-t_0),\ \ 0 < t-t_0 \le 2T\}.$$

Under these conditions, we are dealing with configurations imitating the functioning of M during time long enough for M to stop. As soon as the head stops, it turns into $final$, which expands in both directions due to *rule 0.* This scenario has a positive probability. Therefore it happens somewhere almost sure, whence our process tends to δ(all *final*).

The other direction: Let us assume that M never stops, i.e. continues to function forever if started at the empty tape. Let us take the initial measure concentrated in the configuration "all components are in the empty state" and prove that the resulting distributions cannot contain a stop-head with a positive probability and therefore cannot tend to δ(all *final*). This would be evident if every head functioned alone, never interacting with other heads.

Let us show that in our process every head either functions as if it were alone or disappears. In our construction every head creates its own "territory" marked by left bracket at the left end and by right bracket at the right end. This territory consists of sites which this head has visited. However, this territory may be invaded by another head, which changes the states of cells which it visits, and our head must recognize when it happens. Every time when a head wants to move beyond its territory, it carries the bracket one step further, perhaps, invading another head's territory. In this case, due to *rule 1-left*, it changes the symbol on the tape to the empty one, so its functioning does not differ from functioning of a solitary head on a tape which was empty at the beginning. The crucial question is what happens when some head returns to a place which was its territory, but was invaded by another head. If our head does not notice that the site was invaded and uses symbol written there by another head, it may eventually stop although it would not stop if functioned alone. We must avoid this.

Let us examine the situation in more detail. Since right and left are symmetric, it is sufficient to examine, what happens if some head wants to move right. If it has a right bracket, it means that it is expanding its territory; in this case it can do it due to *rule 1-left* and in doing it it will erase the former tape symbol and write the initial symbol as if it were alone on the tape. If it has no right bracket and its right neighbor has no left bracket, it means that it is moving within its own territory and goes to a place, which has never been invaded and the symbol in the right neighbor was written by itself - see *rule 2-left*. However, if it has no right bracket. but its right neighbor has a left bracket, it means that another head has visited this site. In this case our head commits a suicide, that is turns into a no-head. This happens because, on one hand. due to *rule 2-center*, it is not any more where it was, but, on the other hand, this head does not emerge in its right neighbor cell due to *rule 3-left* and the cell, which it wanted to invade, remains intact. All this assures that every head either moves within its own territory, never visited by other heads, expanding it and imitating the functioning of the original Turing machine with one head, or disappears. Therefore the probability that a stop-head will ever emerge is zero and the process does not tend to $\delta(\text{all } \textit{final})$. Thus lemma 4.1 is proved, whence theorem 4.1 immediately follows.

Another important question is what does it mean that we know or can calculate a number. Several times throughout this course we mentioned that this or that critical value is "unknown". But what does it mean to know a number? Do we know $\sqrt{2}$ or π? Oh, yes, we have special notations for them. We can invent special notations for all the critical values, whose existence we have proved, but it will not make us happy. What we really need is to approximate them with any degree of precision. Let us say that we can *calculate* a number x if there is an algorithm, which for any rational number $y \neq x$ decides, whether y is less or greater than x. (We ignore practical limitations and assume that we have unlimited time and memory.) In this sense we can calculate $\sqrt{2}$. π and many other important numbers. Can we calculate in this sense those critical values, existence of which we proved in this course? The answer is unknown except for a few of them, which we know exactly (for example, the critical value of bond percolation on checkered paper, which equals 1/2). Can we calculate parameters of those non-degenerate invariant measures, existence of which we have proved?

The answer is also unknown except some cases when we know them exactly.

Notes.

4.1. *Exercise.* Let us consider the class of operators $R_\alpha^\beta D$ on $\{0,1\}^U$, where $U = \mathbb{Z}^d$ and D is any operator defined by formula (9) with only one restriction: the number of neighbors $n = 1$. Present an algorithm, which decides for all operators of this class, which of them are ergodic and which are not.

4.2. *Exercise.* Let us consider the class of monotonic deterministic operators $D : \Omega \to \Omega$, where $\Omega = \{0,1\}^U$, where $U = \mathbb{Z}^d$, defined by (9). Present an algorithm to decide, which of these operators are ergodic. (Since every deterministic operator can be interpreted as a random operator, we can apply to them the notion of ergodicity.)

4.3. *Exercise.* Prove that the problem of deciding, which cellular automata have only one invariant measure, is algorithmically unsolvable. This statement is very similar to theorem 4.1, but is not identical with it and need to be proved separately because we don't know whether uniqueness of invariant measure implies ergodicity.

4.4. *Exercise.* Prove that the problem of deciding, which cellular automata have more than two linearly independent invariant measures, is algorithmically unsolvable.

4.5. *Unsolved problem.* Let us consider the class of deterministic operators $D : \Omega \to \Omega$, where $\Omega = \{0,1\}^{\mathbb{Z}}$, defined by (9) (unlike exercise 4.2, without assumption of monotonicity). Is there an algorithm to decide which of them are ergodic?

4.6. *Exercise.* The following statement contradicts theorem 4.1: *There is an algorithm to decide which operators of the class described above are ergodic and which are not.* The following is presented as a proof of this statement. Let us denote A the set of operators described above. The set A is countable because we can represent it as

$$A = A_1 \cup A_2 \cup A_3 \cup \ldots,$$

where A_n is the set of those elements of A for which the set of states of a component is $\{1, \ldots, n\}$. Every operator belonging to A_n is determined by the transition function $f(v|x, y, z)$, where $v, x, y, z \in \{1, \ldots, n\}$. Therefore every A_n is finite, whence A is countable. Now let us represent $A = A_e \cup A_n$, where A_e is the set of ergodic operators and A_n is the set of non-ergodic operators. Both A_e and A_n are infinite. (Prove it.) It is well-known that every infinite subset of a countable set is also countable. Therefore both A_e and A_n are countable. A set is called countable if its elements can be enumerated by natural numbers. Thus we can enumerate A_e and A_n as follows:

$$A_e = \{E_1,\ E_2,\ E_3, \ldots\} \quad \text{and} \quad A_n = \{N_1,\ N_2,\ N_3, \ldots\}.$$

After that we can enumerate all elements of A as follows:

$$A = \{E_1,\ N_1,\ E_2,\ N_2,\ E_3,\ N_3, \ldots\}.$$

After that we can formulate an algorithm to decide whether an element of A is ergodic or not: if its number in this enumeration is odd, it is ergodic, otherwise it is non-ergodic. *We claim to have proved both theorem 4.1 and its negation. Is it possible? If not, what is wrong?*

4.7. *Solved problem.* The problem of deciding which operators are ergodic is also unsolvable when every component has only two states, but the state of i-th component depends on the states of components in the range $[i-n, i+n]$ at the previous time [T-00].

4.8. *Solved problem.* Let us consider all uniform deterministic cellular automata on $\Omega = \{0,1\}^{\mathbb{Z}}$, not only monotonic. Deciding, which of them are 0-eroders, is algorithmically unsolvable [Discr].

4.9. *Unsolved problem.* Let us consider the class of cellular automata $R_\alpha^\beta\, D$, where D is defined by (9) and α and β are any rational numbers in $[0,1]$. Is there an algorithm to decide which of them are ergodic?

5. A general approach to cellular automata

Cellular automata we considered till now were defined for this or that particular purpose. Let us present a general definition of cellular automata and prove some general statements about them. Now U is an arbitrary finite or countable set. For every $i \in U$ there is a finite set S_i of states.[6] Most examples considered in literature have identical S_i for all i, but we don't need to assume it in general. As before, the configuration space Ω is the product of S_i over all $i \in U$. Thin cylinders are defined in the same way as before (4), as well as σ-algebra and the set $\mathcal{M}$ of normed measures. As before, convergence of measures means convergence on all thin cylinders.

Now let us speak about operators. As before, for every $i \in U$ there is a finite set $V(i) \subset U$ called its *neighborhood* and we denote $S_{V(i)} = \prod_{j \in V(i)} S_j$. Elements of $V(i)$ are called neighbors of i. For any set $I \subset U$ we denote its neighborhood $V(I) = \cup_{i \in I} V(i)$. Deterministic cellular automata, that is operators $D : \Omega \to \Omega$, are defined by functions $f_i : S_{V(i)} \to S_i$ as follows:

$$(Dx)_i = f_i(x_{V(i)}) \quad \text{for all} \quad i \in U.$$

Now let us define a random cellular automaton P. Let us call a measure on a product-space a *product-measure* if all its components are independent of each other. The way we define our operators is by far not unique, but it is very natural: for any delta-measure $\delta(x)$ the measure $P\,\delta(x)$ is a product-measure. Let us call the distribution of the i-th component according to this measure the *transitional distribution* and denote it $\theta_i(\cdot|x)$. In fact, the i-th transitional distribution depends only on components of x in the neighborhood of i, so we can write it also as $\theta_i(\cdot|x_{V(i)})$. where $x_{V(i)}$ is restriction of x to $V(i)$. By $\theta_i(y|x)$ we denote the value of $\theta_i(\cdot|x)$ on $y \in S_i$, that is the conditional probability that after application of operator P the i-th component will be in the state y if before its application the neighborhood of i was in the state $x_{V(i)}$. This probability is called *transitional probability.*

Thus we have defined how P acts on all delta-measures. If U is finite, this is sufficient because Ω is finite also and any measure is a linear combination of delta-measures. If U is infinite, measures on Ω generally are not finite linear combinations of delta-measures, but as soon as we concentrate on the value of $P\mu$ on a certain thin

[6] To consider growth models like that in note 2.8 we should allow S_i to be infinite, but then some statements of this chapter would be false, so they should better be considered separately.

cylinder, we may restrict μ to $V(I)$, where I is the support of this thin cylinder. Since I is finite, $V(I)$ is finite also, and this restriction is a linear combination of delta-measures on $V(i)$. Thus we can write an explicit formula for the value of $P\mu$ at an arbitrary thin cylinder:

$$(P\mu)(y_i = b_i,\ i \in I) \ =$$
$$\sum_{a_j, j \in V(I)} \mu(x_i = a_i,\ i \in V(I)) \prod_{i \in I} \theta_i(b_i | a_{V(i)}) \tag{14}$$

for any finite set $I \subset U$ and any $b_i,\ i \in I$. As before, our notations are simplified; for example, the formula $(P\mu)(y_i = b_i, \quad i \in I)$ means the value of the measure $P\mu$ on the set

$$\{y \in \Omega \ : \ y_i = b_i \quad \text{for all} \quad i \in I\}.$$

Any deterministic operator can be considered as a degenerate random operator, which transforms any delta-measure into a delta-measure. Its transition probabilities are

$$\theta_i(y | x_{V(i)}) = \begin{cases} 1 & \text{if } y = f_i(x_{V(i)}), \\ 0 & \text{otherwise.} \end{cases}$$

Any percolation operator $P = R_\alpha\, D$ defined in chapter 2 can be represented in the form (14) by taking $V(i) = \{i + v_1, \ldots, i + v_n\}$ and

$$\theta_i(1|x) = \begin{cases} 0 & \text{if } x_j = 0 \text{ for all } j \in V(i), \\ 1 - \alpha & \text{otherwise.} \end{cases}$$

Any product $R_\alpha^\beta\, D$ is an operator of the type defined above with the same neighborhoods and transition probabilities

$$\theta_i(1|x) = \begin{cases} 1 - \alpha & \text{if } f_i(x_{V(i)}) = 1, \\ \beta & \text{if } f_i(x_{V(i)}) = 0. \end{cases}$$

(The values $\theta_i(0|x)$ equal $1 - \theta_i(1|x)$.) We call a random operator degenerate if at least one of its transition probabilities equals zero and non-degenerate only if all its transition probabilities are strictly positive. For example, all deterministic and percolation operators are degenerate. Lemmas 2.1 and 2.2 remain true in the general setting and can be proved in the same way.

Theorem 5.1. The set $\mathcal{M}$ of normed measures on Ω is compact, which means that from any sequence of elements of $\mathcal{M}$ it is possible to select a subsequence which converges to an element of $\mathcal{M}$.

Proof of theorem 5.1. Since the set of thin cylinders is countable, we can enumerate all of them. Let $C_1,\ C_2,\ C_3,\ \ldots$ be a list of all thin cylinders. Now let us take an arbitrary sequence

$$\mu_i \in \mathcal{M} \tag{15}$$

and prove that it has a converging subsequence ν_i. Let us consider the sequence of $\mu_i(C_1)$, i.e. values of our measures on C_1. These values are real numbers between zero and one, so their sequence has a converging subsequence. So the sequence (15) has a subsequence μ_i^1 whose values on C_1 converge. Let us define $\nu_1 = \mu_1^1$ and consider the sequence of values of

$$\mu_i^1(C_2), \quad i = 2, 3, 4, \ldots \tag{16}$$

Again, this sequence has a converging subsequence, whence the sequence (16) has a subsequence μ_i^2 of measures whose values on C_1 and C_2 converge. Let us define $\nu_2 = \mu_1^2$ and consider values of measures

$$\mu_i^2(C_3), \quad i = 2, 3, 4, \ldots \tag{17}$$

Arguing in the same way, we obtain its subsequence whose values on C_3 converge and so on. Thus we define ν_1, ν_2, ν_3, ..., whose values on all cylinder sets converge. Thus we found a subsequence which converges on all thin cylinders. From lemma 5.1 it converges to a normed measure. *Theorem 5.1 is proved.*

We call a measure μ *invariant* for operator P if $P\,\mu = \mu$. If the limit $\lim_{t\to\infty} P^t\,\mu$ exists, then it is invariant for P. Therefore our task to study ergodicity of cellular automata is closely connected with study of their invariant measures.

Theorem 5.2. Any operator P defined by (14) has at least one invariant measure.

Proof of theorem 5.2. Let us apply our operator P iteratively to an arbitrary initial measure μ. We obtain a sequence of measures $\mu,\ P\mu,\ P^2\mu,\ P^3\mu, \ldots$. Let us form another sequence of measures $\psi_1,\ \psi_2,\ \psi_3,\ \ldots$, where

$$\psi_k \;=\; \frac{1}{k}\,(\mu + \cdots + P^{k-1}\mu), \quad k = 1, 2, 3, \ldots \tag{18}$$

From theorem 5.1 this sequence has a subsequence which converges to some measure ϕ. Let us prove that ϕ is invariant for P. Suppose that it is not, which means that there is a thin cylinder

$$C \;=\; \{y \;:\; y_i = a_i \quad \text{for all} \quad i \in I\}$$

on which ϕ and $P\phi$ have different values:

$$\phi(C) \neq (P\phi)(C). \tag{19}$$

Let us denote $H = |\phi(C) - (P\phi)(C)| > 0$. Let us denote also

$$C_a = \{x : x_i = a_i \quad \text{for all} \quad i \in V(I)\}, \quad \text{where} \quad a \in S_{V(I)}.$$

Using these notations, we can rewrite the formula (14) as

$$(P\mu)(C) \;=\; \sum_a \mu(C_a) \prod_{i\in I} \theta_i(b_i|a). \tag{20}$$

Since the sequence (18) has a subsequence, which converges to ϕ, we can take k so large that

$$|\psi_k(C) - \phi(C)| < H/3$$

and

$$|\psi_k(C_a) - \phi(C_a)| \;<\; \frac{H}{3\prod_{j\in V(I)}|S_j|} \tag{21}$$

for all a, where $|S_j|$ is the cardinality of S_j. Then

$$P\psi_k \;=\; \frac{1}{k}\,(P\mu + \cdots + P^k\mu),$$

whence

$$\psi_k(C) - P\psi_k(C) \;=\; \frac{\mu(C) - (P^k\mu)(C)}{k},$$

which is by modulo not greater than $2/k$. This is less than $H/3$ as soon as we choose $k > 6/H$,

Also let us prove that $|\,(P\phi)(C) - (P\psi_k)(C)\,| < H/3$:

$$|\,(P\phi)(C) - (P\psi_k)(C)\,| =$$

$$\left| \sum_{a_{V(I)}} \phi(x_I = a_I) \prod_{i\in I} \theta_i(b_i|a_{V(i)}) - \sum_{a_{V(I)}} \psi_i(x_I = a_I) \prod_{i\in I} \theta_i(b_i|a_{V(i)}) \right| =$$

$$\left| \sum_{a_{V(I)}} (\phi - \psi_k)(x_I = a_I) \prod_{i\in I} \theta_i(b_i|a_{V(i)}) \right|. \tag{22}$$

Since every transition probability does not exceed one,

$$\prod_{i\in I} \theta_i(b_i|a_{V(i)}) \;\le\; 1.$$

Hence from (21), (22) does not exceed $H/3$. Thus

$$|\,\phi(C) - (P\phi)(C)\,| \le$$

$$|\,\phi(C) - \psi_k(C)\,| + |\,\psi_k(C) - (P\psi_k)(C)\,| + |\,P\psi_k(C) - (P\phi)(C)\,| <$$
$$H/3 + H/3 + H/3 = H.$$

This contradiction shows that our assumption (19) was false. *Theorem 5.2 is proved.*

In fact, we often need a similar, but more general theorem, which can be formulated as follows.

Theorem 5.3. Suppose that we have a non-empty convex compact subset C of $\mathcal{M}$ where a cellular automaton P acts. Suppose that there is $\mu \in \mathcal{M}$ such that $P^t\mu$ belongds to C for all t. Then P has an invariant measure which belongs to C.

This theorem can be proved in the same way as theorem 5.2 and we leave its proof to the reader. Let us use this theorem to say something new about Stavskaya operator described in chapter 2. Let C be the set of measures for which the density of zeros does not exceed, say, 1/3. Then, taking the initial measure concentrated in "all ones" and α such that $27\alpha/(1-27\alpha) < 1/3$ and using theorem 5.3, we conclude that this operator has an invariant measure in C. The same is true if instead of 1/3 we take any positive number.

In the previous chapters we left many plausible facts unproved. To prove them, notions of *monotonicity* are very useful. They are explained, e.g., in chapter 2 of [Discr] in much detail, but in a somewhat formal manner. Essentially these notions confirm our intuitive feelings about order. We already spoke about monotonic deterministic operators and now shall speak about monotonic random operators also. Let us assume that every S_i is ordered (perhaps, partially). For example, if elements of S_i are integer numbers, they may be ordered in the same way as on the number line. We shall use signs $\prec$ and $\succ$ and words *preceeds* and *succeeds* speaking about this order. For example, when $S_i \equiv \{0,1\}$, we typically assume that $0 \prec 1$ which means that zero preceeds one and (which is the same) $1 \succ 0$ which means that one succeeds zero. These relations are a generalization of relations "less" and "greater" among real numbers, but in the present case there may be uncomparable elements. For any $a, b \in S_i$ we assume that if $a \prec b$ and $b \prec a$, then $a = b$. Let us introduce a partial order on Ω by saying that configuration x *preceeds* configuration y or, what is the same, y *succeeeds* x and writing $x \prec y$ or $y \succ x$ if $x_i \leq y_i$ for all $i \in U$. We call a deterministic operator D monotonic if $x \prec y$ implies $Dx \prec Dy$. This definition is consistent with what we said before. Let us say that a measurable set S is *upper* if

$$(x \in S \text{ and } x \prec y) \implies y \in S.$$

Analogously, a set S is *lower* if

$$(y \in S \text{ and } x \prec y) \implies x \in S.$$

It is easy to check that a complement to an upper set is lower and vice versa.

We introduce a partial order on $\mathcal{M}$ by saying that a normed measure μ *preceeds* ν or ν *succeeeds* μ if $\mu(S) \leq \nu(S)$ for any upper S (or $\mu(S) \geq \nu(S)$ for any lower S, which is equivalent). We call an operator P monotonic if $\mu \prec \nu$ implies $P\mu \prec P\nu$. Also let us say that operator P_1 *preceeds* operator P_2 or P_2 *succeeeds* P_1 and write $P_1 \prec P_2$ or $P_2 \succ P_1$ if $P_1 \mu \prec P_2 \mu$ for all measures μ. Notice that there are uncomparable configurations, none of which preceeds the other, for example $(0,1)$ and $(1,0)$, even if all elements of S_i are comparable. Similarly there are uncomparable measures and uncomparable operators. Notice also that all our definitions are consistent: if we consider a deterministic operator as a degenerate random operator, our definitions of monotonicity coincide.

Lemma 5.1. Let us have two product-measures μ, $\nu \in \mathcal{M}$, where $\mu = \prod_i \mu_i$ and $\nu = \prod_i \nu_i$. Then $\mu \prec \nu$ if and only if $\mu_i \prec \nu_i$ for all i.

Proof is easy and we omit it.

Of course, superposition of monotonic operators is monotonic, so to know that a superposition of two operators is monotonic, it is sufficient to check monotonicity of each. How to check monotonicity of an operator?

Proposition 5.1. An operator P defined by (14) is monotonic if and only if all the transition distributions $\theta_i(\cdot|x)$ are monotonic functions of x, that is

$$\Big(x \prec y \Big) \implies \theta_i(\cdot|x) \prec \theta_i(\cdot|y). \tag{23}$$

For example, all percolation operators satisfy this condition and therefore are monotonic.

Proof of proposition 5.1. *In one direction:* suppose that (23) is false, that is there are i, and $y \prec z$ such that $\theta_i(\cdot|y)$ does not preceed $\theta_i(\cdot|z)$. Then $\delta(y) \prec \delta(z)$ serve as those $\mu \prec \nu$ for which $P\mu$ does not preceed $P\nu$ because both are product-measures and the i-th factor of $P\mu$ does not preceed the i-th factor of $P\nu$. From lemma 5.1 this is sufficient. *In the opposite direction* it also follows from lemma 5.1.

Proposition 5.2. Given two operators P_1 and P_2 with one and the same Ω and transition distributions $\theta_i^1(\cdot|x)$ and $\theta_i^2(\cdot|x)$ respectively. Then $P_1 \prec P_2$ if and only if

$$\theta_i^1(\cdot|x) \prec \theta_i^2(\cdot|x) \tag{24}$$

for all $i \in U$ and $x \in \Omega$.

Proof of proposition 5.2. *In one direction:* Let us assume that $\theta_i^1(\cdot|y)$ does not preceed $\theta_i^2(\cdot|y)$ for some $i \in U$ and some $y \in \Omega$ and prove that P_1 does not preceed P_2, that is exists a measure μ such that $P_1\ \mu$ does not preceed $P_2\ \mu$. Let us take $\mu = \delta(y)$. Then both $P_1\mu$ and $P_2\mu$ are product-measures, the i-th factors of which violate condition of lemma 5.1, whence $P_1\mu$ does not preceed $P_2\mu$. *In the opposite direction:* Now assume (24) and prove that $P_1\ \mu \prec P_2\ \mu$ for all μ. It is sufficient to prove this for delta-measures, for which it follows from lemma 5.1. *Proposition 3.2 is proved.*

Starting here we assume that all S_i equal $\{0,1\}$ and apply theory of monotonicity to this case.

Lemma 5.2. Let $S_i \equiv \{0,1\}$. If P is monotonic, then sequences $P^t\delta_0$ and $P^t\delta_1$ converge.

Proof. It is easy to prove by induction that

$$\delta_0 \prec P\delta_0 \prec P^2\delta_0 \prec P^3\delta_0 \prec P^4\delta_0 \dots$$

and

$$\delta_1 \succ P\delta_1 \succ P^2\delta_1 \succ P^3\delta_1 \succ P^4\delta_1 \dots.$$

Indeed, in each case the first inequality is evident because δ_0 preceeds and δ_1 succeeeds any measure and all the other inequalities follow from this. Thus for any upper or lower set C the sequences $P^t\delta_1(C)$ and $P^t\delta_0(C)$ are monotonic and therefore each of them has a limit. Now let us take any thin cylinder C and denote

$$\overline{C} = \{x \mid \exists y \in C : y \prec x\} \quad \text{and} \quad \overline{C}' = \overline{C} \setminus C.$$

It is easy to show that $\overline{C}$ and $\overline{C}'$ are upper sets, so the values of $P^t\delta_0$ and $P^t\delta_1$ at these sets have limits. Therefore their values at C also have limits, which means that these measures have limits. *Lemma 5.2 is proved.*

One example of application of lemma 5.2: the sequence $P^t\ \delta_1$ for the Stavskaya operator P has a limit; since from (8)

$$P^t\ \delta_1\ (x_0 = 0)\ \leq\ \frac{27\alpha}{1-27\alpha} \quad \text{for all} \quad t,$$

the same inequality is true for the limit measure, whence for small values of α the Stavskaya operator has at least two different invariant measures.

Now let us finish the proof of theorem 2.1. It follows from proposition 5.2 that if we have two percolation operators P_1 and P_2 with neighborhoods V_1 and V_2 respectively and $V_1 \subset V_2$, then $P_1 \prec P_2$. This immediately implies the remaining part of theorem 2.1.

Now let us prove lemma 2.4 and thereby finish the proof of theorem 2.3. Let us assume that our operator $P = R_\alpha^\beta\, D$ is not ergodic. From theorem 5.2 we know that P has an invariant measure μ. Let us assume that there is ν such that $P^t\nu$ does not tend to μ. This means that there is a thin cylinder C such that $P^t\nu(C)$ does not tend to $\mu(C)$. This means that there is $H > 0$ such that for any T there is $t \geq T$ such that

$$|(P^t\nu)(C) - \mu(C)| \ \geq \ H. \tag{25}$$

Let us denote I the support of C and m the number of points in I. Now let us choose the initial conditions of the marginals x and y distributed according to μ and ν respectively. Choose T so large that $p_t < H/m$ for all $t \geq T$. Then for any point (v,t) the probability that it is a point of difference does not exceed p_t and therefore for any m points the probability that at least one of them is a point of difference is less than H. Let us denote $E_x(t)$ the event "process x at time t is in C" and $E_y(t)$ the event "process y at time t is in C". The symmetric difference ΔE of these two events is contained in the event "at least one of the points $(v,t), v \in I$ is point of difference". But we know that probability of the latter event is less than H, whence the probability of ΔE also is less than H, whence the difference of probabilities of $E_x(t)$ and $E_y(t)$ also is less than H. But this contradicts (25). This contradiction proves that our assumption about non-ergodicity of P was false. Lemma 2.4 and theorem 2.3 are proved.

Now let us prove theorem 3.4. Since $R_\alpha^\beta \prec R^\beta$,

$$(R_\alpha^\beta\, D)^t\, \delta_0 \ \prec \ (R^\beta\, D)^t\, \delta_0$$

for all t. But for small enough β the density of ones in $(R^\beta\, D)^t\, \delta_0$ does not exceed $1/3$ for all t provided β is small enough. Therefore the density of ones in $(R_\alpha^\beta\, D)^t\, \delta_0$ also does not exceed $1/3$ for all t. Then from theorem 5.3 operator $R_\alpha^\beta\, D$ has an invariant measure μ_0, whose density of ones does not exceed $1/3$. But, since

D is an 1-eroder also, we can use similar arguments to prove that $R_\alpha^\beta\ D$ has an invariant measure μ_1, whose density of zeros does not exceed 1/3. Therefore $\mu_0 \neq \mu_1$. *Theorem 3.4 is proved.*

We have presented some non-degenerate non-ergodic cellular automata, for which $U = \mathbb{Z}^2$, so we may call them two-dimensional. Similar constructions and arguments can be presented for all dimensions greater than one. Are there one-dimensional non-degenerate non-ergodic cellular automata? For a long time it was a common opinion in statistical physics that phase transitions can occur only in systems, whose dimension is greater than one. For example, §152 of Landau and Lifshitz's famous monograph [Lan+Lif] was called "The impossibility of the existence of phases in one-dimensional systems" and an argument of physical nature was presented in support of this impossibility. Another example: "In one dimension bosons do not condense, electrons do not superconduct, ferromagnets do not magnetize, and liquids do not freeze" [Lie+Mat], p. vi.[7] However, all these ideas were formed in dealing with models of equilibrium, which had no time parameter. Cellular automata, besides space, have time. Should it be counted as an additional dimension? The results of modelling [Pet+Pia+Vas] suggested that not. After that the *positive rates conjecture* was proposed by several authors based on various informal considerations. It claimed that all uniform non-degenerate one-dimensional random cellular automata with finite S_i are ergodic. (see, for example, Chapter 4, section 3 of [Liggett], or p. 115 of [Discr] or [Gray]).

However, nothing can substitute a rigorous proof. The systems, which mathematicians consider, may be too complicated to model and much more general than those which arise from physical considerations, and may contradict physical intuition. Now the positive rates conjecture is refuted: P. Gács proposed a non-ergodic non-degenerate uniform one-dimensional system. Like most examples presented in this course, Gács's system actually is an operator acting on $\Omega = S^{\mathbb{Z}}$, which is a composition of a deterministic and a random operator, the random operator turning any state into any other state with a small probability. The main property of the system is that errors do not accumulate, so that the density of components in "wrong" states remains small forever. The system is very complicated and some defects were found in its first version, but now all of them are corrected and an updated version of Gács's construction will be published soon by Journal of Statistical Physics [Gacs-01]. It takes more than two hundred pages

[7] Our note 1.5 indicates in this direction also.

to describe and it needs, although finite, but enormous number of elements in the set S of states of a single component and although positive but very small probability of error. Lately I asked Gács to estimate, at least roughly, these parameters of his construction. He was not sure, but suggested 2^{100} as a rough estimate of the number of states and one divided by a square of this number as a rough estimation of probability of error. Although Gács's result is very important theoretically, these numbers make any practical application very unlikely. It would be interesting to find out, whether such a large number of states and such a small probability of error are really necessary. Also it would be interesting to study various simple one-dimensional constructions having some properties similar to phase transition.

Notes.

5.1. *Exercise.* Given two measures μ and ν on one and the same space, such that $\mu \prec \nu$ and $\nu \prec \mu$. What can you conclude about μ and ν?

5.2. *Exercise.* Where in the proof of lemma 5.1 and theorem 5.1 did we use the condition that all S_i are finite? Do lemma 5.1 and theorem 5.1 remain true if some S_i is infinite?

5.3. *Exercise.* Prove that the operator R_α^β is monotonic if and only if $\alpha + \beta \leq 1$.

5.4. *Exercise.* Prove that the operator $R_\alpha^\beta\, D_{NEC}$ is ergodic as soon as $2/3 < \alpha + \beta < 4/3$.

5.5. *Exercise.* Let us call a random operator P *anti-monotonic* if $\mu \prec \nu \Longrightarrow P\mu \succ P\nu$.
a) Prove that R_α^β is anti-monotonic if and only if $\alpha + \beta \geq 1$.
b) Prove that superposition of two anti-monotonic operators is monotonic.
c) Can an operator be neither monotonic nor anti-monotonic?
d) Prove an analog of proposition 5.1: An operator P defined by (14) is anti-monotonic if and only if

$$\Big(x \prec y \Big) \implies \theta_i(\cdot|x) \succ \theta_i(\cdot|y).$$

e) If an operator is both monotonic and anti-monotonic, what can you say about it?

5.6. *Exercise.* By *coupling* of two or more measures we mean a measure on a product of their spaces, whose marginals are given measures. Prove that, given two measures μ and ν on one and the same space, $\mu \prec \nu$ if and only if there is a coupling of μ and ν, according to which $x \prec y$ a.s., where x and y are configurations in the first and second factor spaces.

5.7. *Unsolved problem.* Does uniqueness of invariant measure imply ergodicity? In other words, is there a non-ergodic cellular automaton, which has one invariant measure?

5.8. *Mean-field approximation.* Let us consider Stavskaya operator in which all points are randomly mixed after every step. In this case one parameter x_t, the density of zeros, is sufficient to describe what happens at every time step. This parameter at all times is determined by the conditions:

$$x_0 = 0, \quad x_{t+1} = \alpha + (1-\alpha)\, x_t^2. \tag{26}$$

Prove that the limit $\lim_{t\to\infty} x_t$ exists and study its behavior depending on parameter α. For which values of α this limit equals 1 and for which values of α this limit is less than 1? Of course, behavior of x_t is different from behavior of the density of zeros at time t in Stavskaya process. However, their qualitative similairy is intriguing: in both cases there is a critical value of α. In physics such approximations of complex processes by simple iterations are widely used and called "mean-field approximations" because they can be interpreted as a substitution of individual particles and their interactions by some uniformly distrubuted mean field described by just one parameter - density. The iteration (26) is only an approximation for the Stavskaya operator, but it is exact for an analogous operator on a special graph known as Bethe graph. Here it is:

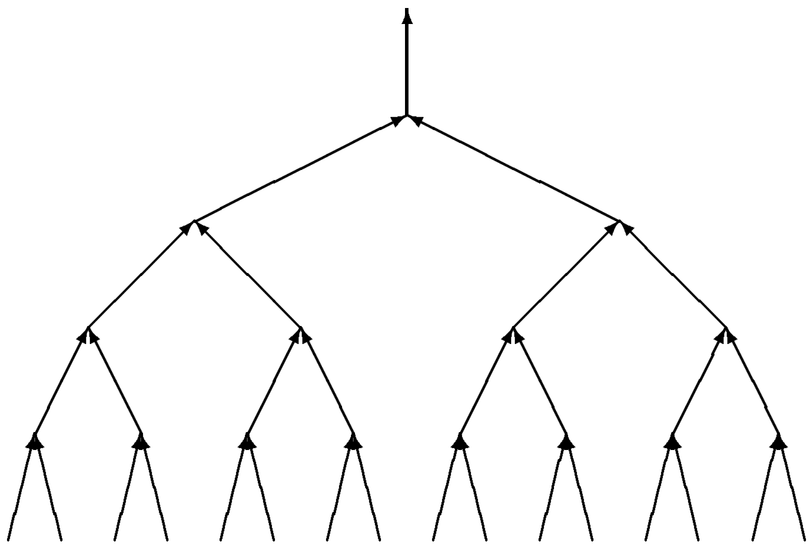

Figure 5.1. Part of Bethe graph, where the mean-field approximation for uniform operators with two neighbors is exact because any product-measure turns into a product-measure.

5.9. *Exercise.* Write a mean-field approximation for the operator $R_\alpha^\beta\, D_{NEC}$ and study for which values of parameters it is ergodic.

Main terms and notations

$\mathbb{Z}^d$ - the d-dimensional integer space, a product of d factors, everyone of which is $\mathbb{Z}$, the set of integer numbers.

$\mathbb{R}^d$ - the d-dimensional real space, a product of d factors, everyone of which is $\mathbb{R}$, the set of real numbers.

Path in a graph - a finite or infinite sequence "vertex-bond-vertex-bond...", where every bond connects those vertices between which it is places in this sequence.

Contour - a path in which the first and last vertices coincide.

U - a finite or countable set, discrete analog of physical space

S_i - set of states of i-th component.

Configuration space - product-space $\Omega = \prod_{i \in U} S_i$. Most attention is given to the case $S_i \equiv \{0,1\}$, that is $\Omega = \{0,1\}^U$.

Configuration - an element of configuration space.

$I_a(x)$ - the set of points v where $x_v = a$.

Thin cylinder - subset of Ω of the form

$$\{x \in \Omega : x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}\}.$$

Support of this thin cylinder is the set $\{i_1, \ldots, i_n\}$.

Normed measure μ on Ω is defined by its values on thin cylinders. "Normed" means $\mu(\Omega) = 1$.

$\mathcal{M}$ - set of normed measures on Ω.

Delta-measure $\delta(x)$ - measure concentrated on configuration x. Measures δ_0 and δ_1 are concentrated on "all zeros" and "all ones".

Product-measure - a measure on a product-space, in which all the marginals are independent.

Cellular automaton - same as linear *operator* $P : \mathcal{M} \to \mathcal{M}$, which transforms any delta-measure $\delta(x)$ into a product-measure, i-th factor of which depends only on $x_{V(i)}$, where $V(i)$ is finite for every i and $x_{V(i)}$ is restriction of x to $V(i)$.

$P\,\mu$ - result of application of operator P to measure μ.

Transition distribution $\theta_i(\cdot|x)$ - distribution of the i-th component according to the measure $P\ \delta(x)$.

Transition probability $\theta_i(y|x)$ - probability that the i-th component equals y according to the measure $P\ \delta(x)$.

Degenerate measure - a measure which equals zero at at least one thin cylinder.

Degenerate cellular automaton - a cellular automaton, at least one transition probability of which is zero.

Superposition PQ of two operators P and Q - an operator, whose action consists of action first Q, then P.

Uniform measure - a measure, which is invariant under space shifts.

Uniform operator - an operator which commutes with space shifts.

Invariant measure: a measure $\mu \in \mathcal{M}$ is called invariant for operator P if $P\,\mu = \mu$.

Ergodicity: Operator P is called *ergodic* if the limit $\lim_{t\to\infty} P^t\mu$ exists and is one and the same for all $\mu \in \mathcal{M}$.

Monotonicity: A deterministic operator D is *monotonic* if $x \prec y$ implies $D\,x \prec D\,y$. A random operator P is *monotonic* if $\mu \prec \nu$ implies $P\,\mu \prec P\,\nu$.

Coupling of two or more measures - a measure on a product-space, whose marginals are given measures. *Coupling* of two or more processes - a process on a product-space, whose marginals are given processes.

x-eroder - a deterministic cellular automaton, for which the configuration x is invariant and which "erodes" all finite deviations from x by turning them into x.

0-eroder - x-eroder for $x =$ "all zeros".

1-eroder - x-eroder for $x =$ "all ones".

Shift of a set S in a linear space at a vector v denoted $S+v$ - the set $\{i+v \mid i \in S\}$.

Vector sum of two sets in a linear space -
$S_1 + S_2 = \{i+j \mid i \in S_1,\ j \in S_2\}$.

Convex set - a set in a linear space, which with any two points a, b contains the segment $[a, b]$.

Convex hull of a set S in a linear space - intersection of all convex sets containing S. It is the "minimal" convex set containing S in the sense that any its convex proper subset does not contain S.

Turing machine - an abstract "machine" proposed by Alan Turing as a formalization of the notion of algorithm.

References

[Ben+Gri] C. Bennett and G. Grinstein. Role of Irreversibility in Stabilizing Complex and Nonergodic Behavior in Locally Interacting Discrete Systems. *Phys. Rev. Letters*, v. 55 (1985), n. 7, pp. 657-660.

[Ber+Sim] P. Berman and J. Simon. Investigations of Fault-Tolerant Networks of Computers *ACM Symp. on Theory of Computing*, 20 (1988), 66-77.

[Bra+Gra] M. Bramson and L. Gray. A useful renormalization argument. *Festschrift for F. Spitzer.* Birkhäuser, Boston, MA.

[Cell] A. Toom. Cellular Automata with Errors: Problems for Students of Probability. *Topics in Contemporary Probability and its Applications.* Ed. J. Laurie Snell. Series *Probability and Stochastics* ed. by Richard Durrett and Mark Pinsky. CRC Press, 1995, pp. 117-157.

[Gacs-01] P. Gács. Reliable cellular automata with self-organization. To appear in *Journal of Stat. Physics*, 2001.

[Galperin] G. Galperin. Homogeneous local monotone operators with memory. Doklady of Soviet Acad. of Sciences, 228, pp. 277-280, 1976.

[Gray] L. F. Gray. The Positive Rates Problem for Attractive Nearest Neighbor Spin Systems on *Z . Z. Wahrscheinlichkeitstheorie verw. Gebiete*, V. 61 (1982), pp. 389-404.

[Grimmett] Geoffrey Grimmett. Percolation. Springer, 1999.

[Discr] Discrete local Markov systems. A. Toom, N. Vasilyev, O. Stavskaya, L. Mityushin, G. Kurdyumov and S. Pirogov. *Stochastic Cellular Systems : ergodicity, memory, morphogenesis.* Ed. by R. Dobrushin, V. Kryukov and A. Toom. Nonlinear Science: theory and applications, Manchester University Press, 1990, pp. 1-182.

[Kurdyumov] G. L. Kurdyumov. An algorithm-theoretic method in studying homogeneous random networks, In: R. L. Dobrushin, V. I. Kryukov, and A. L. Toom (editors). Locally Interacting Systems and Their Application in Biloogy. Lecture Notes in Maqthematics, 653, Springer, pp. 37-55.

[Lan+Lif] L. D. Landau and E. M. Lifshitz. Statistical Physics. (Vol. 5 of *Course of Theoretical Physics.*) 2-d Edition. Pergamon Press, 1969.

[Leb+Mae+Spe] J. L. Lebowitz, C. Maes and E. R. Speer. Statistical mechanics of probabilistic cellular automata. *Journal of Stat. Physics* 59 (1990), 1-2, 117-168.

[Lie+Mat] Mathematical Physics in One Dimension. Exactly Soluble Models of Interacting Particles. A Collection of Reprints with Introductory Text by Elliott H. Lieb and Daniel C. Mattis. N.Y., Academic Press, 1966.

[Liggett] Thomas M. Liggett. Interacting Particle Systems. N.Y., Springer-Verlag, 1985.

[Minsky] Marvin L. Minsky. Computation: finite and infinite machines. Prentice-Hall, 1967.

[Pet+Pia+Vas] M. Petrovskaya, I. Piatetski-Shapiro, and N. Vasilyev. Modelling of voting with random errors. Automatics and Telemechanics, v.10, pp. 103-107, 1969 (in Russian).

[Rockafellar] R. Tyrrell Rockafellar. Convex analysis. Princeton University Press, 1970.

[Sta+Pia] O. Stavskaya and I. Piatetski-Shapiro. On homogeneous nets of spontaneously active elements. *Systems Theory Res.*, v. 20 (1971), pp. 75-88. Originally this article was published in Russian in 1968.

[Tof+Mar] T. Toffoli and N. Margolus. Programmable matter, Concepts and realization. *Physica D* 47 (1991) 263-272.

[T-80] A. Toom. Stable and attractive trajectories in multicomponent systems. *Multicomponent Random Systems*, ed. by R. Dobrushin and Ya. Sinai. Advances in Probability and Related Topics, Dekker, 1980, v. 6, pp. 549-576.

[T-94] A. Toom. On critical phenomena in interacting growth systems. Parts I, II. *Journal of Stat. Physics*, 1994, v. 74, n. 1/2, pp. 91-130.

[T-00] A. Toom. Algorithmical unsolvability of the ergodicity problem for binary cellular automata. *Markov Processes and Related Fields*, v. 6, n. 4, 2000, pp. 569-577.

[Wang] Hao Wang. Popular lectures on mathematical logic. Dover publications, Inc., New York, 1981.