Minimal entropy approximation for cellular automata

Henryk Fukś

Department of Mathematics, Brock University, St. Catharines, ON, L2S 3A1, Canada
E-mail: hfuks@brocku.ca

Received 21 June 2013
Accepted for publication 8 January 2014
Published 14 February 2014

Online at stacks.iop.org/JSTAT/2014/P02009
doi:10.1088/1742-5468/2014/02/P02009

Abstract. We present a method for the construction of approximate orbits of measures under the action of cellular automata which is complementary to the local structure theory. The local structure theory is based on the idea of Bayesian extension, that is, construction of a probability measure consistent with given block probabilities and maximizing entropy. If instead of maximizing entropy one minimizes it, one can develop another method for the construction of approximate orbits, at the heart of which is the iteration of finite-dimensional maps, called minimal entropy maps. We present numerical evidence that the minimal entropy approximation sometimes outperforms the local structure theory in characterizing the properties of cellular automata. The density response curve for elementary CA rule 26 is used to illustrate this claim.

Keywords: cellular automata
1. Introduction

Suppose that one iterates a discrete dynamical system with a local interaction (e.g., cellular automaton), starting from a random initial configuration. If the initial configuration is an infinite binary string with a given probability of occurrence of 1, what is the probability of occurrence of 1 after $n$ iterations? Since the exact answer to this question is usually difficult to obtain, an approximate method for computing such probabilities has been proposed, called the local structure theory. Although in general the local structure theory has very good convergence properties, there are some rare cases when it converges very slowly. In this paper we will explore an alternative approximation which appears to outperform the local structure theory in such cases.

Let $\mathcal{A} = \{0, 1, \ldots, N - 1\}$ be called an alphabet, or a symbol set, and let $X = \mathcal{A}^{\mathbb{Z}}$. Elements of $X$ will be called configurations. A finite sequence of elements of $\mathcal{A}$, $b = b_1b_2, \ldots, b_n$ will be called a block (or word) of length $n$. The set of all blocks of elements of $\mathcal{A}$ of all possible lengths will be denoted by $\mathcal{A}^*$. A Cylinder set generated by the block $b = b_1b_2\ldots b_n$ and anchored at $i$ is defined as

$$[b]_i = \{x \in \mathcal{A}^{\mathbb{Z}} : x_{[i,i+n)} = b\}. \quad (1)$$

The appropriate mathematical description of a distribution of configurations is a probability measure on $X$. Cellular automata (CA) are often considered as maps in the space of such probability measures [1]–[4].

In this paper, we will be interested in shift-invariant probability measures over $X$, or more precisely, in shift-invariant probability measures on the $\sigma$-algebra generated by elementary cylinder sets of $X$. The set of such measures will be denoted by $\mathcal{M}_\sigma(X)$. 

doi:10.1088/1742-5468/2014/02/P02009
Detailed construction of measures in $\mathcal{M}_\sigma(X)$ is described in the review article [5], and the interested reader is advised to consult this reference. Nevertheless, it is not necessary to be familiar with the details of the construction in order to follow the present paper.

The most important feature of any measure $\mu \in \mathcal{M}_\sigma(X)$ is that it is fully determined by measures of cylinder sets $\mu([a])$ for all $a \in \mathcal{A}^*$, which we will denote by $P(a) = \mu([a])$. 

Note that $P(a)$, which we will call block probability, is independent of $i$ due to shift-invariance of the measure $\mu$. Block probability $P(a)$ can be intuitively understood as the probability of occurrence of a given block $a$ in the distribution of configurations.

The following theorem formally states the connection between block probabilities and measures. It is a direct consequence of the Hahn–Kolmogorov extension theorem. For proof the reader can consult [5] and references therein.

**Theorem 1.1.** Let $P : \mathcal{A}^* \rightarrow [0, 1]$ satisfy the conditions

\begin{align}
P(b) &= \sum_{a \in \mathcal{A}} P(ba) = \sum_{a \in \mathcal{G}} P(ab) \quad \forall b \in \mathcal{A}^* , \\
1 &= \sum_{a \in \mathcal{G}} P(a).
\end{align}

Then $P$ uniquely determines shift-invariant probability measure on the $\sigma$-algebra generated by elementary cylinder sets of $X$.

The conditions (3) and (4) are known in the literature as consistency conditions.

Although the set of all block probabilities $\{P(a) : a \in \mathcal{A}^*\}$ is countable, it is still infinite, and in many practical problems, such as computer simulations, it is often possible to keep track of only a finite number of block probabilities. This brings an important question: if we know the probabilities of all blocks of a given length, can we reconstruct the entire measure approximately? One answer to this question is well known and called ‘Bayesian extension’, originally introduced in the context of lattice gases [6, 7]. The approximate measure produced by the Bayesian extension is known as a ‘finite-block measure’ or as ‘Markov process with memory’. The aforementioned review paper [5] discusses details of the Bayesian extension. The main feature of this extension is that, given a finite set of probabilities, it constructs all other block probabilities, satisfying consistency conditions, such that the resulting measure has the maximal entropy.

The Bayesian extension has proved to be a very useful device in statistical physics as well as in the theory of cellular automata. In 1987, Gutowitz et al [8] proposed a generalization of the mean-field theory for cellular automata based on the idea of Bayesian extension. They called it local structure theory. The local structure theory, recently formalized and extended [5], turned out to be a very powerful tool for the characterization of cellular automata.

Given the success of the local structure theory, which is based on the maximal entropy approximation, it seems quite natural to ask how useful the complementary approximation would be, namely the one which minimizes the entropy instead of maximizing it? To the knowledge of the author, no one has ever pursued this idea, and this paper is intended to partially fill this gap.
In what follows, we will investigate the minimal entropy approximation in a configuration space over a binary alphabet, that is, assuming $\mathcal{A} = \{0, 1\}$. Although the ideas presented below can be easily carried over to alphabets of higher cardinality, the binary case is the simplest and the most elegant one, and that is the only reason why we restrict our attention to $\mathcal{A} = \{0, 1\}$.

2. Minimal entropy extension

Before we proceed, let us define $P^{(k)}$ to be the column vector of all probabilities of blocks of length $k$ arranged in lexical order. For example, for $\mathcal{A} = \{0, 1\}$, the first three vectors $P^{(k)}$ are

\begin{align*}
P^{(1)} &= [P(0), P(1)]^T, \\
P^{(2)} &= [P(00), P(01), P(10), P(11)]^T, \\
P^{(3)} &= [P(000), P(001), P(010), P(011), P(100), P(101), P(110), P(111)]^T, \\
&\ldots.
\end{align*}

The entropy of $P^{(k)}$ will be defined as

$$h(P^{(k)}) = -\sum_{b \in \mathcal{A}^k} P(b) \log P(b). \quad (5)$$

Suppose that for a given probability measure we know all block probabilities $P^{(1)}, P^{(2)}, \ldots, P^{(n)}$. We want to construct block probabilities $P^{(n+1)}$ which minimize the entropy $h(P^{(n+1)})$ and are consistent with block probabilities $P^{(1)}, P^{(2)}, \ldots, P^{(n)}$.

In order to do this, we first must remark that not all block probabilities which are components of vectors $P^{(1)}, P^{(2)}, \ldots, P^{(n)}$ are independent, due to consistency conditions. In [5], we demonstrated that for $\mathcal{A} = \{0, 1\}$, only $2^n - 1$ block probabilities are independent. Which ones are declared to be independent, and which ones are treated as dependent, is to some extent arbitrary. One choice of independent probabilities is called short form representation [5]. For the binary alphabet, in the short form representation, block probabilities which have the form $P(0a0)$ are declared to be independent, and the remaining ones are treated as dependent. For example, among the components of $P^{(1)}, P^{(2)}, P^{(3)}$, the independent probabilities are $P(0), P(00), P(000), P(010)$. The remaining ones can be expressed as

\begin{align*}
\begin{bmatrix}
P(01) \\
P(00)
\end{bmatrix}
&= \begin{bmatrix}
P(0) - P(00) \\
P(0) - P(00) - P(010)
\end{bmatrix}, \\
\begin{bmatrix}
P(10) \\
P(11)
\end{bmatrix}
&= \begin{bmatrix}
P(0) - P(00) \\
1 - 3P(0) + 2P(00) + P(010)
\end{bmatrix}, \\
P(1) &= 1 - P(0).
\end{align*}

$$
Coming back to our problem, if we want to construct $P^{(n+1)}$ given $P^{(1)}, P^{(2)}, \ldots, P^{(n)}$, we are free to choose only the values of components of $P^{(n+1)}$ which are of the form $P^{(0a0)}$, where $a \in A^{n-1}$. These probabilities will be denoted by $x_a$, and the remaining ones can be expressed in terms of $x_a$ and probabilities of shorter blocks,

$$
\begin{align*}
P^{(0a0)} &= x_a, \\
P^{(0a1)} &= P^{(0a)} - x_a, \\
P^{(1a0)} &= P^{(a0)} - P^{(0a0)} = P^{(a0)} - x_a, \\
P^{(1a1)} &= P^{(a1)} - P^{(0a1)} = P^{(a1)} - (P^{(0a)} - x_a).
\end{align*}
$$

Let us now consider the following problem: how to choose parameters $x_a$ in order to minimize entropy $h(P^{(n+1)})$? The following theorem provides the answer.

**Theorem 2.1.** Suppose that $\mu$ is a shift-invariant probability measure, and $P^{(b)} = \mu([b])$. Let us define a family of block probabilities $\hat{P}$ such that

$$
\begin{align*}
\hat{P}^{(0a0)} &= \hat{x_a}, \\
\hat{P}^{(0a1)} &= P^{(0a)} - \hat{x_a}, \\
\hat{P}^{(1a0)} &= P^{(a0)} - \hat{x_a}, \\
\hat{P}^{(1a1)} &= P^{(a1)} - (P^{(0a)} - \hat{x_a}),
\end{align*}
$$

where

$$
\begin{align*}
\hat{x}_a &= \begin{cases} 
\max\{0, P^{(0a)} - P^{(a1)}\} & \text{if } |P^{(a1)} - P^{(0a)}| < |P^{(a0)} - P^{(0a)}|, \\
\min\{P^{(0a)}, P^{(a0)}\} & \text{otherwise}.
\end{cases}
\end{align*}
$$

Then

$$
- \sum_{b \in A^{n+1}} \hat{P}^{(b)} \log \hat{P}^{(b)} \leq - \sum_{b \in A^{n+1}} P^{(b)} \log P^{(b)}. 
$$

**Proof.** Let us first notice that $P^{(0a)}$, $P^{(a0)}$, and $P^{(a1)}$ are not independent. Consistency conditions imply that

$$
P^{(0a)} \leq P^{(a1)} + P^{(a0)} \leq 1,
$$

and from there we obtain

$$
P^{(0a)} - P^{(a0)} \leq P^{(a1)} \leq 1 - P^{(a0)}. 
$$

Denoting $\alpha_a = P^{(0a)}$, $\beta_a = P^{(a0)}$, $\delta_a = P^{(a1)}$, this can be written as

$$
\alpha_a - \beta_a \leq \delta_a \leq 1 - \beta_a.
$$

The right-hand side of inequality (10) can be written as

$$
- \sum_{b \in A^{n+1}} P^{(b)} \log P^{(b)} = - \sum_{a \in A^{n-1}} P^{(0a0)} \log P^{(0a0)} + P^{(0a1)} \log P^{(0a1)}
$$

$$
+ P^{(1a0)} \log P^{(1a0)} + P^{(1a1)} \log P^{(1a1)}. 
$$

\doi{10.1088/1742-5468/2014/02/P02009}
Using equation (7) this becomes

\[- \sum_{b \in \mathcal{A}^{n+1}} P(b) \log P(b) = \sum_{a \in \mathcal{A}^{n-1}} H_a(x_a),\]  

(15)

where we define

\[H_a(x_a) = -x_a \log x_a - (\alpha_a - x_a) \log (\alpha_a - x_a)\]

\[- (\beta_a - x_a) \log (\beta_a - x_a) - (\delta_a - (\alpha_a - x_a)) \log (\delta_a - (\alpha_a - x_a)).\]  

(16)

Function \(H_a(x_a)\) is concave, and \(x_a\) can only take values from some closed interval \([x_a, 1]\). For this reason, \(H_a(x_a)\) reaches its minimum at one of the endpoints of the interval. We will show that the minimum occurs precisely at \(x_a = \hat{x}_a\), where \(\hat{x}_a\) is defined in equation (9). First, let us determine the values of the endpoints \(x_{a,1}, x_{a,2}\). In order to do this, note that obviously \(x_a \in [0, 1]\). By consistency conditions,

\[P(a1) = P(0a1) + P(1a1),\]  

(17)

\[P(a0) = P(0a0) + P(1a0),\]  

(18)

\[P(0a) = P(0a0) + P(0a1),\]  

(19)

and therefore

\[P(a1) \geq P(0a1),\]  

(20)

\[P(a0) \geq P(0a0),\]  

(21)

\[P(0a) \geq P(0a1).\]  

(22)

Using equations (7) and the notation \(\alpha_a = P(0a), \beta_a = P(a0), \delta_a = P(a1)\), this becomes

\[\delta_a \geq \alpha_a - x_a,\]  

(23)

\[\beta_a \geq x_a,\]  

(24)

\[\alpha_a \geq x_a.\]  

(25)

Solving the above system of inequalities for \(x_a\) we obtain

\[\alpha_a - \delta_a \leq x_a \leq \min \{\alpha_a, \beta_a\}.\]  

(26)

Since \(x_a \in [0, 1]\), we obtain the following expression for the endpoints of the interval \([x_{a,1}, x_{a,2}]\),

\[x_{a,1} = \max \{0, \alpha_a - \delta_a\}, \quad x_{a,2} = \min \{\alpha_a, \beta_a\}.\]  

(27)

Suppose now that we fix both \(a\) and \(\alpha_a\). Let us consider separately the four cases described in the table below.

| \(\delta_a < \alpha_a\) | \(\beta_a < \alpha_a\) | \(\beta_a \geq \alpha_a\) |
|------------------------|------------------------|------------------------|
| \(x_{a,1} = \alpha_a - \delta_a\) | \(x_{a,1} = \alpha_a - \delta_a\) | \(x_{a,1} = \alpha_a - \delta_a\) |
| \(x_{a,2} = \beta_a\) | \(x_{a,2} = \beta_a\) | \(x_{a,2} = \beta_a\) |

| \(\delta_a \geq \alpha_a\) | \(\beta_a < \alpha_a\) | \(\beta_a \geq \alpha_a\) |
|------------------------|------------------------|------------------------|
| \(x_{a,1} = 0\) | \(x_{a,1} = 0\) | \(x_{a,1} = 0\) |
| \(x_{a,2} = \beta_a\) | \(x_{a,2} = \beta_a\) | \(x_{a,2} = \beta_a\) |

doi:10.1088/1742-5468/2014/02/P02009
We will determine the sign of $H_a(x_a, 1) - H_a(x_a, 2)$. If $H_a(x_a, 1) - H_a(x_a, 2) < 0$, then the minimum of $H_a$ occurs at $x_a, 1$, otherwise at $x_a, 2$. To avoid notational clutter, we will drop the index $a$ from $\alpha, \beta, \delta, a$.

**Case 1:** $\beta < \alpha, \delta < \alpha$.

We have

$$H_a(x_a, 1) - H_a(x_a, 2) = \beta \log \beta + (\alpha - \beta) \log(\alpha - \beta) - (\alpha - \delta) \log(\alpha - \delta) - \delta \log \delta. \quad (28)$$

Defining $f_p(x) = x \log x + (p - x) \log(p - x)$ we can write

$$H_a(x_a, 1) - H_a(x_a, 2) = f_\alpha(\beta) - f_\alpha(\delta). \quad (29)$$

The function $f_\alpha(x)$, defined on interval $x \in (0, \alpha)$, reaches its minimum at $x = \alpha/2$, and has the property $f_\alpha(x) = f_\alpha(\alpha - x)$. This means that $f_\alpha(\delta) > f_\alpha(\beta)$, and thus $H_a(x_a, 1) - H_a(x_a, 2) < 0$, if and only if

$$|\delta - \alpha/2| > |\beta - \alpha/2|. \quad (30)$$

**Case 2:** $\beta < \alpha, \delta \geq \alpha$.

We have

$$H_a(x_a, 1) - H_a(x_a, 2) = -\alpha \log \alpha - (\delta - \alpha) \log(\delta - \alpha)$$

$$+ (\alpha - \beta) \log(\alpha - \beta) + (\beta - \alpha + \delta) \log(\beta - \alpha + \delta)$$

$$= f_\beta(\alpha - \beta) - f_\beta(\alpha). \quad (31)$$

For the same reason as before, $H_a(x_a, 1) - H_a(x_a, 2) < 0$ if and only if

$$|\alpha - \beta| > |\alpha - \delta/2|. \quad (32)$$

**Case 3:** $\beta \geq \alpha, \delta < \alpha$.

We have

$$H_a(x_a, 1) - H_a(x_a, 2) = -(\alpha - \delta) \log(\alpha - \delta)$$

$$- (\beta - \alpha + \delta) \log(\beta - \alpha + \delta) + \alpha \log \alpha + (\beta - \alpha) \log(\beta - \alpha)$$

$$= f_\beta(\alpha) - f_\beta(\alpha - \delta). \quad (33)$$

Again, by the property of $f_\beta$ discussed under Case 1, $H_a(x_a, 1) - H_a(x_a, 2) < 0$ is equivalent to

$$|\alpha - \delta - \beta/2| > |\alpha - \beta/2|. \quad (34)$$

**Case 4:** $\beta \geq \alpha, \delta \geq \alpha$.

doi:10.1088/1742-5468/2014/02/P02009
We have
\[ H_a(x_{a,1}) - H_a(x_{a,2}) = -\beta \log \beta - (\delta - \alpha) \log(\delta - \alpha) \]
\[ + (\beta - \alpha) \log(\beta - \alpha) + \delta \log \delta \]
\[ = g(\delta) - g(\beta), \quad (35) \]
where \( g(y) = y \log y - (y - \alpha) \log(y - \alpha) \). Since for \( y \in (0, \alpha) \)
\[ g'(y) = \log y - \log(y - \alpha) > 0, \quad (36) \]
h\((y)\) is increasing in \((0, \alpha)\). This means that \( g(\delta) > g(\beta) \), or equivalently \( H_a(x_{a,1}) - H_a(x_{a,2}) < 0 \), is satisfied if and only if
\[ \delta < \beta. \quad (37) \]

We obtained four inequalities \((30), (32), (34)\) and \((37)\) for the four cases. We plot in
figure 1 solutions of these four inequalities in \(\beta - \delta\) space using different colors for each case.
One can, however, combine all four cases and describe them by one simple inequality taking
into account the fact that only values of \((\beta, \delta)\) marked by vertical hatching are possible, due
to condition \((13)\). This simple inequality combining all four cases (subject to condition
\((13)\)) is
\[ |\delta - \alpha| < |\beta - \alpha|, \quad (38) \]
as one can easily verify graphically by comparing figures 1 and 2.

To summarize our findings, we demonstrated that the minimum of \( H_a \) occurs at \( \hat{x}_a \),
where
\[ \hat{x}_a = \begin{cases} x_{a,1}, & \text{if } |\delta_a - \alpha_a| < |\beta_a - \alpha_a|, \\ x_{a,2}, & \text{otherwise}, \end{cases} \quad (39) \]
and where \( x_{a,1} \) and \( x_{a,2} \) are defined in equation \((27)\). This is precisely equation \((9)\), and
theorem \(2.1\) then follows directly.
3. Minimal entropy approximation for measures

Using theorem 2.1, we can now construct an approximation of a probability measure complementary to the Bayesian approximation. Let us define

\[
\Upsilon(\alpha, \beta, \delta) = \begin{cases} 
\max \{0, \alpha - \delta\} & \text{if } |\delta - \alpha| < |\beta - \alpha|, \\
\min \{\alpha, \beta\} & \text{otherwise}
\end{cases}
\] (40)

and

\[
\Upsilon_{0,0}(\alpha, \beta, \delta) = \Upsilon(\alpha, \beta, \delta),
\]

(41)
\[
\Upsilon_{0,1}(\alpha, \beta, \delta) = \alpha - \Upsilon(\alpha, \beta, \delta),
\]

(42)
\[
\Upsilon_{1,0}(\alpha, \beta, \delta) = \beta - \Upsilon(\alpha, \beta, \delta),
\]

(43)
\[
\Upsilon_{1,1}(\alpha, \beta, \delta) = \delta - \alpha + \Upsilon(\alpha, \beta, \delta).
\]

(44)

Using this notation and equation (8), one can now express probabilities of \((k + 1)\)-blocks by probabilities of \(k\)-blocks, by writing

\[
P(a_1 a_2 \ldots a_{k+1}) \approx \Upsilon_{a_1,a_{k+1}} \left( P(0a_2 \ldots a_k), P(a_2 \ldots a_k 0), P(a_2 \ldots a_k 1) \right).
\]

(45)

The above approximation will be called minimal entropy approximation. We can, of course, repeat this process, and approximate \((k + 2)\)-block probabilities by \((k + 1)\)-block probabilities, and, by applying the approximation again, \((k + 1)\)-block probabilities by \(k\)-block probabilities. For a given integer \(k\), by recursive application of the minimal entropy approximation, any block probability of length \(p > k\) can be expressed by probabilities of length \(k\). The following definition states this more formally.

**Definition 3.1.** Let \(\mu \in \mathcal{M}_\sigma(X)\) be a measure with associated block probabilities \(P : \mathcal{A}^* \rightarrow [0, 1], P(b) = \mu([b]_i)\) for all \(i \in \mathbb{Z}\) and \(b \in \mathcal{A}^*\). For \(k > 0\), define \(P : \mathcal{A}^* \rightarrow [0, 1]\)
Minimal entropy approximation for cellular automata

recursively so that

\[ \hat{P}(a_1a_2 \ldots a_p) = \begin{cases} P(a_1a_2 \ldots a_p) & \text{if } p \leq k, \\ \Upsilon_{a_1,a_p}(\hat{P}(0a_2 \ldots a_{p-1}), \hat{P}(a_2 \ldots a_{p-1}0), \hat{P}(a_2 \ldots a_{p-1}1)) & \text{if } p > k. \end{cases} \] (46)

The above block probabilities determine a shift-invariant probability measure \( \hat{\mu}^{(k)} \in \mathcal{M}_\sigma(X) \), to be called minimal entropy approximation of \( \mu \) of order \( k \).

The proof that block probabilities \( \hat{P} \) indeed satisfy consistency conditions required by theorem 1.1 is an immediate consequence of the definition of \( \hat{P} \).

It is intuitively clear that as the order of the minimal entropy approximations increases, the quality of the approximation should increase too. The following proposition formalizes this observation.

**Proposition 3.1.** The sequence of \( k \)th order minimal entropy approximations of \( \mu \in \mathcal{M}_\sigma(X) \) weakly converges to \( \mu \) as \( k \to \infty \).

**Proof.** Let \( n > 0 \), \( b \in A^n \) and let us denote \( \hat{P}_k(b) = \hat{\mu}^{(k)}([b]_0) \), \( P(b) = \mu([b]_0) \). Since for \( k \geq n \hat{P}_k(b) = P(b) \), we obviously have \( \lim_{k \to \infty} \hat{P}_k(b) = P(b) \). Since cylinder sets constitute the convergence determining class for measures in \( \mathcal{M}_\sigma(X) \), convergence of block probabilities is equivalent to weak convergence. This leads to the conclusion that \( \hat{\mu}^{(k)} \Rightarrow \mu \).

Note that the proof of the above proposition does not rely on the details of construction of the minimal entropy approximation, but only on the fact that the minimal entropy measure is determined by probabilities of blocks of finite length. Indeed, basically the same proof can be given in the case of the maximal entropy approximation used in the local structure theory \([8, 5]\). It should also be noted that proposition 3.1 does not guarantee uniformity of convergence with respect to \( n \).

Measures \( \mu \) for which \( \hat{\mu}^{(k)} = \mu \) will be called \( k \)th order measures of minimal entropy. The set of such measures over \( X \) will be denoted by \( \mathcal{M}^{(k)}_{\text{ME}}(X) \). Obviously these measures are shift-invariant, \( \mathcal{M}^{(k)}_{\text{ME}}(X) \subset \mathcal{M}^{(k)}_\sigma(X) \). Finally, we need to emphasize that the minimization of entropy performed in the construction of \( \hat{\mu}^{(k)} \) is sequential, meaning that we first construct probabilities of blocks of length \( k + 1 \) minimizing entropy of \( P^{k+1} \), then we construct probabilities of blocks of length \( k + 2 \) consistent with \( P^{k+1} \) and minimizing entropy of \( P^{k+2} \), etc. This means that that for a given \( \mu \), \( k > 0 \) and \( m > k + 1 \), \( \hat{\mu}^{(k)} \) will not necessarily be the measure having the lowest possible entropy of \( P^m \) among measures which agree with \( \mu \) on probabilities of blocks of length up to \( k \).

4. Orbits of measures under the action of cellular automata

Cellular automata (CA) are dynamical systems characterized by discreteness in space and time. They can be viewed as cells in a regular lattice updated synchronously according to a local interaction rule, where the state of each cell is restricted to a finite set of allowed
values. In a probabilistic cellular automaton, lattice sites simultaneously change states with probabilities depending on states of local neighbors.

For the purpose of this paper, we will define probabilistic cellular automata as maps in the space of probability measures. Let \( w : A \times A^{2r+1} \rightarrow [0, 1] \), whose values are denoted by \( w(a|b) \) for \( a \in A, b \in A^{2r+1} \), satisfying \( \sum_{a \in A} w(a|b) = 1 \), be called local transition function of radius \( r \), and its values will be called local transition probabilities. A probabilistic cellular automaton with the local transition function \( w \) is a map \( F : \mathcal{M}(X) \rightarrow \mathcal{M}(X) \) defined as

\[
(F\mu)([b]_i) = \sum_{a \in A^{[b]+2r}} w(b|a)\mu([a]_{i-r}) \quad \text{for all } i \in \mathbb{Z}, b \in A^*,
\]

where we define

\[
w(b|a) = \prod_{j=1}^{[b]} w(b_j|a_j,a_{j+1} \cdots a_{j+2r}).
\]

For shift-invariant measures, one can drop the indices \( i \) and \( i - r \) in equation (47), as block probabilities are independent of the position of the block.

When the function \( w \) takes values in the set \( \{0, 1\} \), the corresponding cellular automaton is called a deterministic CA. Deterministic cellular automata with \( A = \{0, 1\} \) and \( r = 1 \) are called elementary cellular automata.

For any probabilistic measure \( \mu \in \mathcal{M}_\sigma(X) \), we define the orbit of \( \mu \) under \( F \) as

\[
\{F^n\mu\}_{n=0}^\infty.
\]

Excluding trivial cases, computing the orbit of a measure under a given CA explicitly encounters serious difficulties. Even if we are not interested in the complete measure, but only in a probability of a particular block \( b \) after \( n \) iterations of the rule, computing this probability by recursive application of equation (47) requires an increasing computational effort as \( n \) increases, as we have to sum over the exponentially growing set of blocks. In order to avoid this difficulty, Gutowitz et al proposed a method for constructing approximations of orbits based on the minimal entropy approximation [8, 5], known as the local structure theory for CA. Here we will propose a similar method for approximating orbits based on the minimal entropy approximation.

Let us first define the entropy minimizing operator of order \( k \), denoted by \( \Psi^{(k)} \), to be a map from \( \mathcal{M}_\sigma(X) \) to \( \mathcal{M}_{\text{ME}}^{(k)}(X) \) such that

\[
\Psi^{(k)}\mu = \hat{\mu}^{(k)},
\]

where \( \hat{\mu}^{(k)} \) is the measure defined in definition 3.1. Note that the operator \( \Psi^{(k)} \) is idempotent, that is, \( \Psi^{(k)}\Psi^{(k)}\mu = \Psi^{(k)}\mu \). This allows us to construct the approximate orbit of a measure \( \mu \) under the action of \( F \) by simply replacing \( F \) by \( \Psi^{(k)}F\Psi^{(k)} \). The sequence

\[
\left\{ \left( \Psi^{(k)}F\Psi^{(k)} \right)^n\mu \right\}_{n=0}^\infty
\]

will be called the minimal entropy approximation of level \( k \) of the exact orbit \( \{F^n\mu\}_{n=0}^\infty \). Note that all terms of this sequence are measures of minimal entropy, thus the entire approximate orbit lies in \( \mathcal{M}_{\text{ME}}^{(k)}(X) \).
Minimal entropy approximation for cellular automata

Just like for the local structure approximation, the minimal entropy approximation approximates the actual orbit increasingly well as $k$ increases. In fact, we will prove that every point of the approximate orbit weakly converges to the corresponding point of the exact orbit.

**Proposition 4.1.** Let $k$ be a positive integer and $b \in \mathcal{A}^*$. If $k \geq |b| + 2r$, then

$$F\mu([b]) = F\Psi^{(k)}\mu([b]) = \Psi^{(k)}F\mu([b]).$$  \hfill (52)

**Proof.** To prove it, note that $\mu([a]) = \hat{\mu}^{(k)}([a])$ for all blocks $a$ of length up to $k$. The first equality of (52) can be written as

$$\sum_{a \in \mathcal{A}^{(|b|+2r)}} w(b|a)\mu([a]) = \sum_{a \in \mathcal{A}^{(|b|+2r)}} w(b|a)\hat{\mu}^{(k)}([a]).$$  \hfill (53)

The equality holds when $|a| \leq k$, that is, $|b| + 2r \leq k$.

The second equality of (52) is a result of the fact that the $\Psi^{(k)}$ operator only modifies probabilities of blocks of length greater than $k$. Since $k \geq |b| + 2r$, we have $|b| < k$ and therefore $F\mu([b]) = \Psi^{(k)}F\mu([b])$. \hfill $\square$

Now let us note that $F^n$ can be viewed as a cellular automaton rule of radius $nr$, thus when $k \geq |b| + 2nr$, we have $F^n\mu([b]) = F^n\Psi^{(k)}\mu([b])$. We can insert an arbitrary number of $\Psi^{(k)}$ operators on the right-hand side anywhere we want, and nothing will change, because $\Psi^{(k)}$ does not modify relevant block probabilities. This yields an immediate corollary.

**Corollary 4.1.** Let $k$ and $n$ be positive integers and $b \in \mathcal{A}^*$. If $k \geq |b| + 2nr$, then

$$F^n\mu([b]) = \left(\Psi^{(k)}F\Psi^{(k)}\right)^n\mu([b]).$$

This means that for a given $n$, measures of cylinder sets in the approximate measure $(\Psi^{(k)}F\Psi^{(k)})^n\mu$ coincide with measures of cylinder sets in $F^n\mu$ for sufficiently large $k$. Because cylinder sets constitute convergence determining class for measures, we obtain the following result.

**Theorem 4.1.** Let $F$ be a cellular automaton, $\mu \in \mathcal{M}_\sigma(X)$ be a shift-invariant measure, and $\nu^{(k)}_n$ be a minimal entropy approximation of order $k$ of the measure $F^n\mu$, i.e., $\nu^{(k)}_n = (\Psi^{(k)}F\Psi^{(k)})^n\mu$. Then for any positive integer $n$, $\nu^{(k)}_n \Rightarrow F^n\mu$ as $k \to \infty$.

Just as in the case of proposition 3.1, it should be noted that the proof of the above theorem does not depend on the details of the construction of the operator $\Psi^{(k)}$. In fact, any idempotent operator which does not modify probabilities of blocks of length up to $k$ and produces consistent probabilities of blocks longer than $k$ would yield exactly the same theorem.
5. Minimal entropy maps

Minimal entropy maps can be entirely described by specifying a finite number of block probabilities. We will use this feature to constructs a finite-dimensional map which approximates the action of a CA rule on shift-invariant measures. If \( \nu_n^{(k)} = (\Psi^{(k)} F \Psi^{(k)})^n \mu \), then \( \nu_n^{(k)} \) satisfies the recurrence equation

\[
\nu_{n+1}^{(k)} = \Psi^{(k)} F \Psi^{(k)} \nu_n^{(k)}. \tag{54}
\]

On both sides of this equation we have measures in \( \mathfrak{M}_{\text{ME}}^{(k)}(X) \), and these are completely determined by probabilities of blocks of length \( k \). If \( |b| = k \), we obtain

\[
\nu_{n+1}^{(k)}([b]) = \Psi^{(k)} F \Psi^{(k)} \nu_n^{(k)}([b]), \tag{55}
\]

and, since \( \Psi^{(k)} \) does not modify probabilities of blocks of length \( k \), this simplifies to

\[
\nu_{n+1}^{(k)}([b]) = F \Psi^{(k)} \nu_n^{(k)}([b]). \tag{56}
\]

By the definition of \( F \),

\[
\nu_{n+1}^{(k)}([b]) = \sum_{a \in \mathcal{A}^{[b]+2r}} w(b|a) \left( \Psi^{(k)} \nu_n^{(k)}([a]) \right). \tag{57}
\]

To simplify the notation, let us define \( Q_n(b) = \nu_n^{(k)}([b]) \), and, consistent with definition in equation (46), \( \hat{Q}_n(a) = (\Psi^{(k)} \nu_n^{(k)})([a]) \). Then we can rewrite the previous equation to take the form

\[
Q_{n+1}(b) = \sum_{a \in \mathcal{A}^{[b]+2r}} w(b|a) \hat{Q}_n(a). \tag{58}
\]

Note that by equation (46), \( \hat{Q}_n(a) \) depends only on probabilities of blocks of length \( k \). If we thus arrange \( Q_n(b) \) for all \( b \in \mathcal{A}^k \) in lexicographical order to form a vector \( Q_n \), we will obtain

\[
Q_{n+1} = U^{(k)}(Q_n), \tag{59}
\]

where \( U^{(k)} : [0,1]^{|\mathcal{A}|^k} \to [0,1]^{|\mathcal{A}|^k} \) has components defined by equation (58). We will call this map an entropy minimizing map of order \( k \).

6. Example: elementary CA rule 26

As an example, consider a deterministic binary rule of radius 1, defined by

\[
\begin{align*}
w(1|000) &= 0, & w(1|001) &= 1, & w(1|010) &= 0, & w(1|011) &= 1, \\
w(1|100) &= 1, & w(1|101) &= 0, & w(1|110) &= 0, & w(1|111) &= 0.
\end{align*} \tag{60}
\]

According to the Wolfram numbering scheme for CA rules [9], this is so-called elementary CA rule with code number 26. Suppose that that we now we wish to construct the minimal
entropy map of order 2 for this rule. Let $P_n(b) = F^n\mu([b])$. Using equation (47) we obtain for $r = 1$, $|b| = 3$

$$P_{n+1}(b) = \sum_{a \in A^5} w(a|b)P_n(a).$$

(61)

Using the definition of $w(a|b)$ given in equation (48) and the transition probabilities given in equation (60) we obtain

$$P_{n+1}(00) = P_n(0000) + P_n(0101) + P_n(1010) + P_n(1101) + P_n(1110) + P_n(1111),$$
$$P_{n+1}(01) = P_n(0001) + P_n(0100) + P_n(1011) + P_n(1110),$$
$$P_{n+1}(10) = P_n(0010) + P_n(0110) + P_n(0111) + P_n(1000),$$
$$P_{n+1}(11) = P_n(0011) + P_n(1001).$$

(62)

This set of equations describes the exact relationship between block probabilities at step $n + 1$ and block probabilities at step $n$. Note that 3-block probabilities at step $n + 1$ are given in terms of 5-blocks probabilities at step $n$, thus it is not possible to iterate these equations.

The minimal entropy map of order 2 (equation (59)) can be obtained by simply replacing $P$ by $Q$ and placing the operator $\hat{\cdot}$ over probabilities on the right-hand side of equation (62). This yields

$$Q_{n+1}(00) = \hat{Q}_n(0000) + \hat{Q}_n(0101) + \hat{Q}_n(1010) + \hat{Q}_n(1101) + \hat{Q}_n(1110) + \hat{Q}_n(1111),$$
$$Q_{n+1}(01) = \hat{Q}_n(0001) + \hat{Q}_n(0100) + \hat{Q}_n(1011) + \hat{Q}_n(1100),$$
$$Q_{n+1}(10) = \hat{Q}_n(0010) + \hat{Q}_n(0110) + \hat{Q}_n(0111) + \hat{Q}_n(1000),$$
$$Q_{n+1}(11) = \hat{Q}_n(0011) + \hat{Q}_n(1001).$$

(63)

Using equation (46) with $k = 2$, one can express $\hat{Q}_n(a_1a_2a_3a_4)$ in terms of 2-block probabilities. For example,

$$\hat{Q}_n(0000) = \Upsilon(\hat{Q}_n(000), \hat{Q}_n(000), \hat{Q}_n(001)) = \hat{Q}_n(000)$$
$$= \Upsilon(\hat{Q}_n(00), \hat{Q}_n(00), \hat{Q}_n(01)) = \hat{Q}_n(00) = Q_n(00).$$

(64)

Similarly one obtains

$$\hat{Q}_n(0101) = Q_n(01), \quad \hat{Q}_n(1010) = Q_n(10), \quad \hat{Q}_n(1111) = Q_n(11).$$

(65)

All other $\hat{Q}_n(a_1a_2a_3a_4)$ are equal to 0. This simplifies equation (63) to

$$Q_{n+1}(00) = Q_n(00) + Q_n(01) + Q_n(10) + Q_n(11),$$
$$Q_{n+1}(01) = 0,$$
$$Q_{n+1}(10) = 0,$$
$$Q_{n+1}(11) = 0.$$
When $k = 3$, we follow the same procedure as for the $k = 2$ case discussed above. If we write equation (58) for all possible $b \in A^3$, we will have on the left-hand sides eight block probabilities $Q(b_1b_2b_3)$, thus the resulting minimal entropy map will be 8-dimensional,

$$
Q_{n+1}(000) = \hat{Q}_n(00000) + \hat{Q}_n(01010) + \hat{Q}_n(10101) + \hat{Q}_n(11010) + \hat{Q}_n(11101) + \hat{Q}_n(11111),
$$

$$
Q_{n+1}(001) = \hat{Q}_n(00001) + \hat{Q}_n(01011) + \hat{Q}_n(10100) + \hat{Q}_n(11011) + \hat{Q}_n(11100),
$$

$$
Q_{n+1}(010) = \hat{Q}_n(00010) + \hat{Q}_n(01000) + \hat{Q}_n(10110) + \hat{Q}_n(10111) + \hat{Q}_n(11000),
$$

$$
Q_{n+1}(011) = \hat{Q}_n(00011) + \hat{Q}_n(01001) + \hat{Q}_n(11001),
$$

$$
Q_{n+1}(100) = \hat{Q}_n(00100) + \hat{Q}_n(01101) + \hat{Q}_n(11010) + \hat{Q}_n(11111) + \hat{Q}_n(10000),
$$

$$
Q_{n+1}(101) = \hat{Q}_n(00101) + \hat{Q}_n(01100) + \hat{Q}_n(10011) + \hat{Q}_n(10001),
$$

$$
Q_{n+1}(110) = \hat{Q}_n(00110) + \hat{Q}_n(00111) + \hat{Q}_n(10010),
$$

$$
Q_{n+1}(111) = \hat{Q}_n(10011).
$$

(67)

On the right-hand side, we have 32 block probabilities which have to be expressed in terms of 3-block probabilities by using equation (46) with $k = 3$. Some of these will simplify to a single 3-block probability, e.g.,

$$
\hat{Q}_n(00000) = \Upsilon(\hat{Q}_n(0000), \hat{Q}_n(0000), \hat{Q}_n(0001)) = \hat{Q}_n(0000) = Q_n(000).
$$

(68)

Others, in general, will not simplify, and will have to be expressed by nested $\Upsilon$ functions, for example

$$
\hat{Q}_n(00100) = \Upsilon\left(\Upsilon(Q_n(001), Q_n(010), Q_n(011)), \Upsilon(Q_n(010), Q_n(100), Q_n(101))\right).
$$

(69)

Once we express all $\hat{Q}_n(a_1a_2a_3a_4a_5)$ in equation (67) by 3-block probabilities $Q_n(a_1a_2a_3)$, we obtain a map $[0, 1]^8 \to [0, 1]^8$. We omit explicit formulas for this map due to its complexity. One should stress, however, that only four components of this map are independent, and that by exploiting consistency conditions for block probabilities it is possible to reduce this map to $[0, 1]^4 \to [0, 1]^4$. We refer the interested reader to [5], where we explained how to perform such a reduction for local structure maps (the same method can used for minimal entropy maps).

Just for the sake of comparison, let us also write the local structure map of order 3 for rule 26. It can be obtained from equation (67) by replacing $\hat{Q}$ with $\tilde{Q}$,

$$
Q_{n+1}(000) = \tilde{Q}_n(00000) + \tilde{Q}_n(01010) + \tilde{Q}_n(10101) + \tilde{Q}_n(11010) + \tilde{Q}_n(11101) + \tilde{Q}_n(11111),
$$

$$
Q_{n+1}(001) = \tilde{Q}_n(00001) + \tilde{Q}_n(01011) + \tilde{Q}_n(10100) + \tilde{Q}_n(11011) + \tilde{Q}_n(11100),
$$

$$
Q_{n+1}(010) = \tilde{Q}_n(00010) + \tilde{Q}_n(01000) + \tilde{Q}_n(10110) + \tilde{Q}_n(10111) + \tilde{Q}_n(11000),
$$

$$
Q_{n+1}(011) = \tilde{Q}_n(00011) + \tilde{Q}_n(01001) + \tilde{Q}_n(11001),
$$

$$
Q_{n+1}(100) = \tilde{Q}_n(00100) + \tilde{Q}_n(01101) + \tilde{Q}_n(01110) + \tilde{Q}_n(01111) + \tilde{Q}_n(10000),
$$

$$
Q_{n+1}(101) = \tilde{Q}_n(00101) + \tilde{Q}_n(01100) + \tilde{Q}_n(10011) + \tilde{Q}_n(10001),
$$

$$
Q_{n+1}(110) = \tilde{Q}_n(00110) + \tilde{Q}_n(00111) + \tilde{Q}_n(10010),
$$

$$
Q_{n+1}(111) = \tilde{Q}_n(10011).
$$

(70)
Minimal entropy approximation for cellular automata

where
\[
\hat{Q}_n(a_1a_2a_3a_4a_5) = \frac{Q_n(a_1a_2a_3)Q_n(a_2a_3a_4)Q_n(a_3a_4a_5)}{(Q_n(a_2a_3a_4a_5) + Q_n(a_2a_3a_4a_5)) (Q_n(a_3a_4a_5) + Q_n(a_2a_3a_4a_5) + Q_n(a_2a_3a_4a_5))}.
\] (71)

Both minimal entropy maps and local structure maps become rather complicated when \(k\) increases. Because of high dimensionality and strong nonlinearity, it is difficult to perform standard stability analysis for these maps. It is, however, rather straightforward to write a computer program which constructs and iterates them.

7. Experimental results

As we have already mentioned, orbits of minimal entropy maps approximate orbits of measures under cellular automata rules. By iterating the minimal entropy map, we can obtain approximate \(P_n(a)\), that is, probability of occurrence of block \(a\) after \(n\) iterations of a given cellular automata rule. How good is this approximation, and is it any better than the local structure approximation?

In order to shed some light on this question, we considered the following problem. Suppose that the initial measure is a Bernoulli measure \(\mu_p\), so that
\[
\mu_p([a]) = P_0([a]) = p^j (1-p)^{|a|-j},
\] (72)
where \(j\) is the number of ones in \(a\), \(|a|-j\) is the number of zeros in \(a\), and \(p \in [0,1]\). The probability of occurrence of \(a\) after \(n\) iterations is then given by
\[
P_n(a) = (F^n \mu_p)([a]).
\] (73)
The expected value of a given cell after the \(n\)th iteration of the rule, to be denoted \(\rho_n\), is given by
\[
\rho_n = 1 \cdot P_n(1) + 0 \cdot P_n(0) = P_n(1).
\] (74)
We will call \(\rho_n\) a density of ones at time \(n\). The density can be estimated numerically by starting with an array of \(N\) sites and setting each one of them independently to 1 or 0 with probability \(p\) or \(1-p\), respectively. We then iterate rule \(F_n\) times (using periodic boundary conditions) and count how many cells are in state 1. The count divided by \(N\) serves as a numerical estimate of \(\rho_n\).

One can also estimate \(\rho_n\) by iterating the \(k\)th order minimal entropy map \(n\) times starting from the initial conditions given by equation (72), that is, \(Q_0(a) = P_0(a)\). Then we compute \(Q_n(1)\) by using the consistency conditions,
\[
Q_n(1) = \sum_{a \in \{0,1\}^{k-1}} Q_n(a1),
\] (75)
and \(Q_n(1)\) is used as an approximation of \(\rho_n\), to be called the \(k\)th order minimal entropy approximation of \(\rho_n\). An analogous approximation using the local structure map will be called the \(k\)th order local structure approximation of \(\rho_n\). An interesting question is now how \(\rho_n\) depends on \(\rho_0\). A plot of \(\rho_n\) versus \(\rho_0\) is called a density response curve.
How well can the local structure theory approximate the density response curve for elementary CA? The answer is ‘remarkably well’. For a vast majority of elementary CA rules, if one plots the density response curves using ‘experimental’ data (that is, data obtained by direct computer simulation) as well as curves obtained by using the local structure approximation, one finds that as the order of the approximation increases, the local structure curves become closer and closer to the ‘experimental curves’. A typical example is shown in figure 3(a), using the elementary CA rule with Wolfram number 108. One can see that already for \( k = 5 \), the local structure theory approximates the density curve exceptionally well. It is interesting to note that the minimal entropy approximation works equally well in this case.

The above does not mean, of course, that we will observe the same phenomenon for all elementary CA rules. In fact, there exist some remarkable exceptions, for which the local structure theory converges so slowly that up to the order \( k = 7 \) one does not see any improvement in the approximation. The most interesting of those is rule 26, discussed in the previous section.

Figure 4(a) shows the experimental density response curve for rule 26 together with density curves obtained by iterations of local structure maps up to order 7. The latter are horizontal straight lines, very far from the experimental curve. One can say, therefore,
Minimal entropy approximation for cellular automata

Figure 4. Density response curves for rule 26 for $t = 10^3$ obtained by iteration of local structure maps (a) and minimal entropy maps (b). Labels LST $k$ and ME $k$ indicate the order of the approximation $k$ for, respectively, the local structure and the minimal entropy approximation.

that the local structure fails to predict the correct shape of the density curve, at least for $k \leq 7$.

Figure 4(b), on the other hand, shows the experimental curve together with curves obtained by iterating minimal entropy maps. One can clearly see that these curves approximate the experimental curve much better, and that the accuracy of the approximation increases with $k$. One can thus say that the minimal entropy approximation outperforms the local structure approximation in this case.

The observations made in this section can be interpreted as follows. Since, as remarked in the note under the theorem 4.1, the weak convergence of approximating measures does not rely on the details of their construction, one expects that both the local structure approximation and the minimal entropy approximation should produce better and better approximations of density response curves as $k$ increases. In fact, any self-consistent scheme for approximating probabilities of blocks longer than $k$ by probabilities of blocks of length $k$ will have that property. Nevertheless, the choice of the scheme does seem to matter if one wants *fast* convergence. Local structure theory is usually good in that respect, yet there are some cases when it seems to converge very slowly, such as for rule 26. In such cases, the minimal entropy approximation appears to be a better choice. This of course does not mean that it is necessarily the best choice—other approximation schemes...
with even better convergence properties may exist. One hopes that this paper stimulates further research on this topic.

8. Conclusions

We introduced the notion of the minimal entropy approximation of probability measures over binary bisequences. Minimal entropy approximation can be viewed as an opposite of Bayesian approximation, which maximizes entropy. We then demonstrated how the minimal entropy approximation can be used to construct approximations of orbits of measures under the action of deterministic or probabilistic cellular automata. Such approximate orbits can be fully characterized by orbits of finite-dimensional maps, which we call minimal entropy maps. While points of approximate orbits of measures obtained by iterating minimal entropy maps weakly converge to corresponding points of the exact orbits, just as in the case of approximate orbits of local structure theory, there are cases when the minimal entropy approximation works better than the local structure approximation. This is the case for elementary CA rule 26, for which the local structure theory fails in predicting the correct shape of the density response curve for $k \leq 7$. The minimal entropy approximation yields fairly accurate predictions for the density response curve of rule 26, starting with $k = 3$.

An interesting question is why is the minimal entropy approximation better than the maximal entropy approximation in the case of rule 26? One could naively think that this is because the time evolution of rule 26 is somewhat more ‘ordered’ than for other rules. It is, however, not true: there are other rules for which the spatiotemporal patterns are even more ‘ordered’ than for rule 26, yet both maximal and minimal entropy approximations seem to work for them equally well. In order to probe this issue further, one will need to find more examples of rules for which the minimal entropy approximation outperforms the local structure theory. A natural way to go beyond elementary CA rules considered here is to search for such examples among either probabilistic CA rules of radius 1, or deterministic CA rules of radius greater than 1. Both possibilities are currently investigated by the author.

Acknowledgments

The author acknowledges partial financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant. Some calculations on which this work is based were made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.

References

[1] Kurka P and Maass A, Limit sets of cellular automata associated to probability measures, 2000 J. Stat. Phys. 100 1031
[2] Kurka P, On the measure attractor of a cellular automaton, 2005 Discrete Contin. Dyn. Syst. suppl. vol. 524 doi:10.1088/1742-5468/2014/02/P02009
Minimal entropy approximation for cellular automata

[3] Pivato M, *Ergodic theory of cellular automata*, 2009 Encyclopedia of Complexity and System Science ed R A Meyers (Berlin: Springer)

[4] Formenti E and Kůrka P, *Dynamics of cellular automata in non-compact spaces*, 2009 Encyclopedia of Complexity and System Science ed R A Meyers (Berlin: Springer)

[5] Fukš H, *Construction of local structure maps for cellular automata*, 2013 J. Cellular Autom 7 455

[6] Brascamp H J, *Equilibrium states for a one dimensional lattice gas*, 1971 Commun. Math. Phys. 21 56

[7] Fannes M and Verbeure A, *On solvable models in classical lattice systems*, 1984 Commun. Math. Phys. 96 115

[8] Gutowtiz H A, Victor J D and Knight B W, *Local structure theory for cellular automata*, 1987 Physica D 28 118

[9] Wolfram S, 1994 *Cellular Automata and Complexity: Collected Papers* (Reading, MA: Addison-Wesley)