Concave Programming Upper Bounds on the Capacity of 2-D Constraints*

Ido Tal  Ron M. Roth
Computer Science Department,
Technion, Haifa 32000, Israel.
Email: idotal@ieee.org, ronny@cs.technion.ac.il

Abstract—The capacity of 1-D constraints is given by the entropy of a corresponding stationary maxentropic Markov chain. Namely, the entropy is maximized over a set of probability distributions, which is defined by some linear requirements. In this paper, certain aspects of this characterization are extended to 2-D constraints. The result is a method for calculating an upper bound on the capacity of 2-D constraints.

The key steps are: The maxentropic stationary probability distribution on square configurations is considered. A set of linear equalities and inequalities is derived from this stationarity. The result is a concave program, which can be easily solved numerically. Our method improves upon previous upper bounds for the capacity of the 2-D “no independent bits” constraint, as well as certain 2-D RLL constraints.

I. INTRODUCTION

Let \( \Sigma \) be a finite alphabet. A one-dimensional (1-D) constraint is a set \( S \) of words over \( \Sigma \). For the \( S \) to be called a 1-D constraint, there must exist an edge-labeled graph \( G \) with the following property: a word \( w = w_1w_2 \ldots w_n \) is in \( S \) iff there exists a path in \( G \) for which the successive edge labels are \( w_1, w_2, \ldots, w_n \) (see [1]).

A two dimensional (2-D) constraint over \( \Sigma \) is a generalization of a 1-D constraint; it is a set \( S \) of rectangular configurations over \( \Sigma \) and is defined through a pair of vertex-labeled graphs \( (G_{row}, G_{col}) \), where \( G_{row} = (V,E_{row},L) \) and \( G_{col} = (V,E_{col},L) \). Namely, both graphs share the same vertex set and the same vertex labeling function \( L : V \rightarrow \Sigma \). The constraint \( S = S(G_{row}, G_{col}) \) consists of all finite rectangular configurations \( \{(u_{i,j}) \} \) over \( \Sigma \) with the following property: Let \( \Lambda \) be the rectangular index set of \( \{(u_{i,j}) \} \). There exists a configuration \( (u_{i,j}) \) over the vertex set \( V \) such that (a) for each \( (i,j) \in \Lambda \) we have \( u_{i,j} = L(u_{i,j}) \); (b) each row in \( (u_{i,j}) \) is a path in \( G_{row} \); (c) each column in \( (u_{i,j}) \) is a path in \( G_{col} \). Examples of 2-D constraints include the square constraint [2], 2-D runlength-limited (RLL) constraints [3], 2-D symmetric runlength-limited (SRL) constraints [4], and the “no isolated bits” constraint [5].

Let \( S \) be a given 2-D constraint over a finite alphabet \( \Sigma \). Denote by \( \Sigma^{M \times N} \) the set of \( M \times N \) configurations over \( \Sigma \), and let

\[
S_{M,N} = \Sigma \cap \Sigma^{M \times N}, \quad S_{M} = \Sigma \cap \Sigma^{M \times M}.
\]

The capacity of \( S \) is equal to

\[
\text{cap}(S) = \lim_{M \to \infty} \frac{1}{M^2} \cdot \log_2 |S_M|.
\]

In this paper, we show a method for calculating an upper bound on \( \text{cap}(S) \). Two other methods of calculating an upper bound on the capacity of a 2-D constraint are the following: The first method is the so called “stripe method,” in which we fix a positive integer \( N \), and bound \( \text{cap}(S) \) by

\[
\text{cap}(S) \leq \lim_{M \to \infty} \frac{1}{M \cdot N} \cdot \log_2 |S_{M,N}|.
\]

Namely, we consider only strips of width \( N \), and essentially get a 1-D constraint (since we may regard each of the possible row values as a letter in an auxiliary alphabet). The RHS of (2) is easily calculated for modest values of \( N \): Let \( G \) be the edge-labeled graph corresponding to the 1-D constraint, and let \( A_G \) be the adjacency matrix of \( G \). Denote by \( \lambda(A_G) \) the Perron eigenvalue of \( A_G \). By \([1 \S 3.2]\), the RHS of (2) is equal to \( \lambda(A_G) \). The second method for upper-bounding \( \text{cap}(S) \) is the generalization presented by Forchhammer and Justesen [6] to the method of Calkin and Wilf [7].

The capacity of a given 1-D constraint is known to be equal to the value of an optimization program, where the optimization is on the entropy of a certain stationary Markov chain, and is carried out over the conditional probabilities of that chain (see \([1 \S 3.2.3]\)). We try to extend certain aspects of this characterization of capacity to 2-D constraints. What results is a (generally non-tight) upper bound on \( \text{cap}(S) \).

The structure of this paper is as follows. In Section II we set up some notation. Then, in Section III we show the existence of a certain stationary random variable taking values on \( S_M \) and having entropy approaching the capacity of \( S \), as \( M \to \infty \). We then consider a relatively small sub-configuration of that random variable, and denote it by \( X^{(M)} \).

The section concludes with an upper bound on the capacity of \( S \), which is a function of the probability distribution of \( X^{(M)} \). In Section IV we derive a set of linear equations which hold on the probability distribution of \( X^{(M)} \). In Section V we argue as follows: The bound derived in Section III is a function of the probability distribution of \( X^{(M)} \), which we do not know how to calculate; however, by Section IV we know that this probability distribution is subject to a set of linear requirements. Thus, we formalize an optimization problem, where the unknown probability distribution is replaced by

* This work was supported by grant No. 2002197 from the United-States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.
a set of variables, subject to the above-mentioned linear requirements. The maximum of this optimization problem is an upper bound on the capacity of $S$. We then show that this optimization problem is easily solved, since it is an instance of convex programming. In Section VII we show our computational results. Finally, in Section VII we present an asymptotic analysis of our method.

We note at this point that although this paper deals with 2-D constraints, our method can be easily generalized to higher dimensions as well.

II. Notation

This section is devoted to setting up some notation.

A. Index sets and configurations

Denote the set of integers by $\mathbb{Z}$. A (2-D) index set $U \subseteq \mathbb{Z}^2$ is a set of integer pairs. A 2-D configuration over $\Sigma$ with an index set $U$ is a function $w : U \rightarrow \Sigma$. We denote such a configuration as $w = (w_{i,j})_{(i,j)\in U}$, where for all $(i, j) \in U$, we have that $w_{i,j} \in \Sigma$. In this paper, index sets will always be denoted by upper-case Greek letters or upper-case Roman letters in the sans-serif font. Since many of our configurations will be $M \times N$, we have set aside special notation for their index sets; let

$$B_{M,N} = \{(i, j) : 0 \leq i < M, \quad 0 \leq j < N\}.$$ 

Also, denote

$$B_M = B_{M,M} = \{(i, j) : 0 \leq i, j < M\}.$$ 

For indices $\alpha, \beta$ we denote the shifting of $U$ by $(\alpha, \beta)$ as

$$\sigma_{\alpha,\beta}(U) = \{(i + \alpha, j + \beta) : (i, j) \in U\}.$$ 

Moreover, by abuse of notation, let $\sigma_{\alpha,\beta}(w)$ be the shifted configuration (with index set $\sigma(U)$):

$$\sigma_{\alpha,\beta}(w)_{i + \alpha, j + \beta} = w_{i,j}.$$ 

For a configuration $w$ with index set $U$, and an index set $V \subseteq U$, denote the restriction of $w$ to $V$ by $w[V] = (w[V]_{i,j})_{(i,j)\in V}$; namely,

$$w[V]_{i,j} = w_{i,j}, \quad \text{where} \quad (i,j) \in V.$$ 

We denote the restriction of $S$ to $U$ by $S[U]$: 

$$S[U] = \{w : \text{there exists } w' \in S \text{ such that } w'[U] = w\}. \quad (3)$$

B. Strict total order

A strict total order $\prec$ is a relation on $\mathbb{Z}^2 \times \mathbb{Z}^2$, satisfying the following conditions for all $(i_1, j_1), (i_2, j_2), (i_3, j_3) \in \mathbb{Z}^2$.

- If $(i_1, j_1) \neq (i_2, j_2)$, then either $(i_1, j_1) \prec (i_2, j_2)$ or $(i_2, j_2) \prec (i_1, j_1)$, but not both.
- If $(i_1, j_1) = (i_2, j_2)$, then neither $(i_1, j_1) \prec (i_2, j_2)$ nor $(i_2, j_2) \prec (i_1, j_1)$.
- If $(i_1, j_1) \prec (i_2, j_2)$ and $(i_2, j_2) \prec (i_3, j_3)$, then $(i_1, j_1) \prec (i_3, j_3)$.

For $(i, j) \in \mathbb{Z}^2$, define $T_{i,j}^{(\prec)}$ as all the indexes preceding $(i, j)$. Namely,

$$T_{i,j}^{(\prec)} = \{(i', j') \in \mathbb{Z}^2 : (i', j') \prec (i, j)\}.$$ 

C. Entropy

Let $X$ and $Y$ be two random variables. Denote

$$p_x = \text{Prob}(X = x).$$

and

$$p_{y|x} = \text{Prob}(X = x, Y = y) / \text{Prob}(X = x).$$

The entropy of $X$ is denoted by $H(X)$ and is equal to

$$H(X) = \sum_x p_x \log p_x,$$

where the sum is on all $x$ for which $\text{Prob}(X = x)$ is positive. Similarly, we define the conditional entropy $H(Y|X)$ as

$$H(Y|X) = \sum_x p_x \sum_y p_{y|x} \log p_{y|x},$$

where we sum on all $x$ for which $p_x$ is positive and all $y$ for which $p_{y|x}$ is positive.

III. A PRELIMINARY UPPER BOUND ON $\text{cap}(S)$

Let $M$ be a positive integer and let $W$ be a random variable taking values on $S_M$. We say that $W$ is stationary if for all $U \subseteq B_M$, all $\alpha, \beta \in \mathbb{Z}$ such that $\sigma_{\alpha,\beta}(U) \subseteq B_M$, and all $w' \in S[U]$, we have that

$$\text{Prob}(W[U] = w') = \text{Prob}(W[\sigma_{\alpha,\beta}(U)] = \sigma_{\alpha,\beta}(w')).$$

The following is a corollary of [8, Theorem 1.4]. The proof is given in the Appendix.

Theorem I: There exists a series of random variables $W(M)_{M=1}^{\infty}$ with the following properties: (i) Each $W(M)$ takes values on $S_M$. (ii) The probability distribution of $W(M)$ is stationary. (iii) The normalized entropy of $W(M)$ approaches $\text{cap}(S)$,

$$\text{cap}(S) = \lim_{M \rightarrow \infty} \frac{1}{M^2} \cdot H(W(M)). \quad (4)$$

We now proceed towards deriving Lemma 2 below, which gives an upper bound on $\text{cap}(S)$, and makes use of the stationarity property. We note in advance that this bound is not actually meant to be calculated. Thus, its utility will be made clear in the following sections. In order to enhance the exposition, we accompany the derivation with two running examples.

Running Example I: Define the lexicographic order $\prec_{\text{lex}}$ as follows: $(i_1, j_1) \prec_{\text{lex}} (i_2, j_2)$ iff

- $i_1 < i_2$, or
- $(i_1 = i_2$ and $j_1 < j_2)$.

Running Example II: Define the “interleaved raster scan” order $\prec_{\text{irs}}$ as follows: $(i_1, j_1) \prec_{\text{irs}} (i_2, j_2)$ iff

- $i_1 \equiv 0 \pmod{2}$ and $i_2 \equiv 1 \pmod{2}$, or
- $i_1 \equiv i_2 \pmod{2}$ and $i_1 < i_2$, or
- $i_1 = i_2$ and $j_1 < j_2$.

(See Figure 1 for both examples.)

For the rest of this section, fix positive integers $r$ and $s$, and define the index set

$$\Lambda = B_{r,s}.$$
We will refer to $\Lambda$ as “the patch.” The bound we derive in Lemma 2 will be a function of the following:
- the strict total order $\prec$,
- the integers $r$ and $s$, which determine the order $r \times s$ of the patch $\Lambda$,
- an integer $c$, which will denote the number of “colors” we encounter,
- a coloring function $f : \mathbb{Z}^2 \to \{1, 2, \ldots, c\}$, mapping each point in $\mathbb{Z}^2$ to one of $c$ colors,
- $c$ indexes, $(a_\gamma, b_\gamma)\gamma=1^c$, such that for all $1 \leq \gamma \leq c$,
  $$(a_\gamma, b_\gamma) \in \Lambda$$
  (namely, each color $\gamma$ has a designated point in the patch, which may or may not be of color $\gamma$).

The function $f$ must satisfy two requirements, which we now elaborate on. Our first requirement is: for all $1 \leq \gamma \leq c$,
$$\lim_{M \to \infty} \frac{|\{(i,j) \in B_M : f(i,j) = \gamma\}|}{M^2} = \frac{1}{c}. \tag{5}$$

Namely, as the orders of $W^{(M)}$ tend to infinity, each color is equally likely. Our second requirement is as follows: there exist index sets $\Psi_1, \Psi_2, \ldots, \Psi_c \subseteq \Lambda$ such that for all indexes $(i,j) \in \mathbb{Z}^2$,
$$\sigma_{i',j'}(\Psi_\gamma) = \Gamma_{i,j}^{(\gamma)} \cap \sigma_{i',j'}(\Lambda), \tag{6}$$
where $\gamma = f(i,j)$, $i' = a_\gamma - i$, and $j' = b_\gamma - j$. Namely, let $(i,j)$ be such that $f(i,j) = \gamma$, and shift $\Lambda$ such that $(a_\gamma, b_\gamma)$ is shifted to $(i,j)$. Now, consider the set of all indexes in the shifted $\Lambda$ which precede $(i,j)$; this set must be equal to the correspondingly shifted $\Psi_\gamma$.

**Running Example I:** Take $r = 4$ and $s = 7$ as the patch orders. Let the number of colors be $c = 1$. Thus, we must define $f = f_{\text{lex}}$ as follows: for all $(i,j) \in \mathbb{Z}^2$, $f_{\text{lex}}(i,j) = 1$. Take the point corresponding to the single color as $(a_1 = 3, b_1 = 5)$. See also Figure 2(a).

**Running Example II:** As in the previous example, take $r = 3$ and $s = 5$ as the patch orders. Let the number of colors be $c = 2$. Define $f = f_{\text{irs}}$ as follows:
$$f_{\text{irs}}(i,j) = \begin{cases} 1 & i \equiv 0 \pmod{2} \\ 2 & i \equiv 1 \pmod{2} \end{cases}.$$
where hereafter in the proof, $\gamma = f(i,j) = a_i - i$, and $j' = b_j - j$. Define $\partial = B_M \setminus \bar{\partial}$. Note that since $r$ and $s$ are constant, and $\Psi_1, \Psi_2, \ldots, \Psi_c \subseteq \Lambda$, then

$$\frac{|\partial|}{M^2} = O(1/M).$$

Thus, on the one hand, we have

$$\frac{1}{M^2} \sum_{(i,j) \in \partial} H(W_{i,j} | W[T_{i,j} \cap B_M]) \leq \log_2 |\Sigma| \cdot O(1/M).$$

On the other hand, from (6) and (7) we have that for all $(i,j) \in \bar{\partial}$,

$$\sigma_{i', j'}(\Psi_i) \subseteq T_{i,j} \cap B_M.$$ 

Hence, since conditioning reduces entropy [3, Theorem 2.6.5],

$$\frac{1}{M^2} \sum_{(i,j) \in \partial} H(W_{i,j} | W[T_{i,j} \cap B_M]) \leq \frac{1}{M^2} \sum_{(i,j) \in \partial} H(W_{i,j} | W[\sigma_{i', j'}(\Psi_i)]) \leq \frac{1}{M^2} \sum_{(i,j) \in \partial} H(W[(i,j) \cup \sigma_{i', j'}(\Psi_i)] | W[\sigma_{i', j'}(\Psi_i)]) = \frac{1}{M^2} \sum_{(i,j) \in \partial} H(Y_{i,j} | Z_{i,j}).$$

where the last step follows from the stationarity of $W^{(M)}$. Recalling (5), the proof follows.

The following is a simple corollary of Lemma [2]

**Corollary 3:** Let $(W^{(M)})_{M=1}^{\infty}$ be as in Theorem 1 and define

$$X^{(M)} = W^{(M)}[\Lambda].$$

Fix positive integers $r$ and $s$. Let $\ell$ be a positive integer, and let $(\rho(k))_{k=1}^{\ell}$ be non-negative reals such that $\sum_{k=1}^{\ell} \rho(k) = 1$. For every $1 \leq k \leq \ell$, let $\gamma(k), c(k), f(k), (\Psi(k))_{c=1}^{c(k)}$, and $(\alpha(k)^c, b(k)^c)_{c=1}^{c(k)}$ be given. Also, for $1 \leq \gamma \leq c(k)$, let

$$Y_{\gamma}^{(k)} = \{(a_{\gamma}^{(k)}, b_{\gamma}^{(k)})\} \cup \Psi_{\gamma}^{(k)}.$$ 

Define

$$Y_{\gamma}^{(k)} = X^{(M)}[\Psi_{\gamma}^{(k)}]$$

(note that $Y_{\gamma}^{(k)}$ and $Z_{\gamma}^{(k)}$ are functions of $M$). Then,

$$\text{cap}(\Sigma) \leq \limsup_{M \to \infty} \sum_{k=1}^{\ell} \rho(k) c(k) \sum_{\gamma=1}^{c(k)} H(Y_{\gamma}^{(k)} | Z_{\gamma}^{(k)}).$$

Corollary 3 is the most general way we have found to state our results. This generality will indeed help us later on. However, almost none of the intuition is lost if the reader has in mind the much simpler case of

$$\ell = 1, \quad \rho(1) = 1, \quad c(1) = 1, \quad \gamma(1) = \gamma_{\text{lex}}, \quad (a_{1}^{(1)}, b_{1}^{(1)}) = (r-1, t), \quad \text{and} \quad \Psi_{1}^{(1)} = \Lambda \cap T_{(a_{1}^{(1)}, b_{1}^{(1)})}, \quad (8)$$

where $0 \leq t < s$. This simpler case was demonstrated in Running Example 1.

**IV. Linear requirements**

Recall that $X^{(M)} = W^{(M)}[\Lambda]$ is an $r \times s$ sub-configuration of $W^{(M)}$, and thus stationary as well. In this section, we formulate a set of linear requirements (equalities and inequalities) on the probability distribution of $X^{(M)}$. For the rest of this section, let $M$ be fixed and let $X$ be shorthand for $X^{(M)}$.

**A. Linear requirements from stationarity**

In this subsection, we formulate a set of linear requirements that follow from the stationarity of $X^{(M)}$. Let $x \in \Sigma[\Lambda]$ be a realization of $X$. Denote

$$p_x = \text{Prob}(X = x).$$

We start with the trivial requirements. Obviously, we must have for all $x \in \Sigma[\Lambda]$ that

$$p_x \geq 0.$$ 

Also,

$$\sum_{x \in \Sigma[\Lambda]} p_x = 1.$$ 

Next, we show how we can use stationarity to get more linear equations on $(p_x)_{x \in \Sigma[\Lambda]}$. Let

$$A' = \{(i,j) : 0 \leq i < r - 1, \quad 0 \leq j < s\}.$$ 

For $x' \in \Sigma[A']$ we must have by stationarity that

$$\text{Prob}(X[A'] = x') = \text{Prob}(X[\sigma_{1,0}(A')] = \sigma_{1,0}(x')).$$

As a concrete example, suppose that $r = s = 3$. We claim that

$$\text{Prob}(X = 1 0 0 0 0 1) = \text{Prob}(X = 1 0 0 0 0 1),$$

where $*$ denotes “don’t care”.

Both the left-hand and right-hand sides of (9) are marginalizations of $(p_x)_{x}$. Thus, we get a set of linear equations on $(p_x)_{x}$, namely, for all $x' \in \Sigma[A']$,

$$\sum_{x : x[A'] = x'} p_x = \sum_{x : x[\sigma_{1,0}(A')] = \sigma_{1,0}(x')} p_x.$$ 

To get more equations, we now apply the same rational horizontally, instead of vertically. Let

$$A'' = \{(i,j) : 0 \leq i < r, \quad 0 \leq j < s - 1\}.$$ 

for all $x'' \in \Sigma[A'']$,

$$\sum_{x : x[A''] = x''} p_x = \sum_{x : x[\sigma_{0,1}(A'')] = \sigma_{0,1}(x'')} p_x.$$
B. Linear equations from reflection, transposition, and complementation

We now show that if $S$ is reflection, transposition, or complementation invariant (defined below), then we can derive yet more linear equations.

Define $v_M(\cdot)(h_M(\cdot))$ as the vertical (horizontal) reflection of a rectangular configuration with $M$ rows (columns). Namely,

$$(v_M(w))_{i,j} = w_{M-1-i,j}, \quad \text{and} \quad (h_M(w))_{i,j} = w_{i,M-1-j}. $$

Define $\tau$ as the transposition of a configuration. Namely,

$$\tau(w)_{i,j} = w_{j,i}. $$

For $\Sigma = \{0,1\}$, denote by $\text{comp}(w)$ the bitwise complement of a configuration $w$. Namely,

$$\text{comp}(w)_{i,j} = \begin{cases} 1 & \text{if } w_{i,j} = 0 \\ 0 & \text{otherwise}. \end{cases} $$

We state three similar lemmas, and prove the first. The proof of the other two is similar.

**Lemma 4:** Suppose that $S$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in S \iff h_M(w) \in S \iff v_M(w) \in S. $$

Then, w.l.o.g., the probability distribution of $W$ is such that for all $w \in S_M$,

$$\text{Prob}(W = w) = \text{Prob}(W = h_M(w)) = \text{Prob}(W = v_M(w)). \tag{10} $$

**Lemma 5:** Suppose that $S$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in S \iff \tau(w) \in S. $$

Then, w.l.o.g., $W$ is such that for all $w \in S_M$,

$$\text{Prob}(W = w) = \text{Prob}(W = \tau(w)). \tag{11} $$

**Lemma 6:** Suppose that $\Sigma = \{0,1\}$ and $S$ is such that for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in S \iff \text{comp}(w) \in S. $$

Then, w.l.o.g., $W$ is such that for all $w \in S_M$,

$$\text{Prob}(W = w) = \text{Prob}(W = \text{comp}(w)). \tag{12} $$

**Proof of Lemma 5.** Let $h$ and $v$ be shorthand for $h_M$ and $v_M$, respectively. For $M$ fixed, we define a new random variable $W^{\text{new}}$ taking values on $S_M$, with the following distribution: for all $w \in S_M$,

$$\text{Prob}(W^{\text{new}} = w) = \frac{1}{4} \sum_{w' \in \{w,h(w),v(w),h(v(w))\}} \text{Prob}(W = w'). $$

Since $h(h(w)) = v(v(w)) = w$ and $h(v(w)) = v(h(w))$ we get that (10) holds for $W^{\text{new}}$. Moreover, by the concavity of the entropy function,

$$H(W) \leq H(W^{\text{new}}). $$

Thus, the properties defined in Theorem 1 hold for $W^{\text{new}}$. ■

If the condition of Lemma 4 holds, then we get the following equations by stationarity. For all $x \in \Sigma|\Lambda$,

$$p_x = p_{v(x)} = p_{h(x)}. $$

If the condition of Lemma 5 holds, then the following holds by stationarity. Assume w.l.o.g. that $r \leq s$, and let

$$\hat{A} = \{(i,j) : 0 \leq i,j < r\}. $$

For all $x \in \Sigma|\hat{A}$,

$$\sum_{x : x|\hat{A} = \chi} p_x = \sum_{x : x|\hat{A} = \tau(\chi)} p_x. $$

If the condition of Lemma 6 holds, then we get the following equations by stationarity. For all $x \in \Sigma|\Lambda$,

$$p_x = p_{\text{comp}(x)}. $$

V. AN UPPER BOUND ON $\text{cap}(S)$

For the rest of this section, let $r$, $s$, $\ell$, $\rho$, $(\rho,\kappa)$, $f$, $(\gamma,\rho)$, $\Psi$, $(\gamma,\rho)$, and $(\gamma,\rho)$ be given as in Corollary 3. Recall from Corollary 3 that we are interested in $H(Y(\gamma,\rho)|Z(\gamma,\rho))$, in order to bound $\text{cap}(S)$ from above.

As a first step, we fix $M$ and express $H(Y(\gamma,\rho)|Z(\gamma,\rho))$ in terms of the probabilities $(p_x)_\gamma$ of the random variable $X(M)$. For given $1 \leq k \leq \ell$ and $1 \leq \gamma \leq c(k)$, let

$$y = \sum_{x \in \Sigma|\Lambda : x|\Lambda = \chi} p_x = \sum_{x \in \Sigma|\Lambda : x|\Lambda = \tau(\chi)} p_x. $$

be realizations of $Y(\gamma,\rho)$ and $Z(\gamma,\rho)$, respectively. Let

$$p_{\gamma,y} = \text{Prob}(Y(\gamma,\rho) = y) \quad \text{and} \quad p_{\gamma,z} = \text{Prob}(Z(\gamma,\rho) = z) $$

$(p_{\gamma,y}$ and $p_{\gamma,z}$ are functions of $M)$. From here onward, let $p_y$ and $p_z$ be shorthand for $p_{\gamma,y}$ and $p_{\gamma,z}$, respectively. Both $p_y$ and $p_z$ are marginalizations of $(p_x)_\gamma$, namely,

$$p_y = \sum_{x : x|\Lambda = y} p_x, \quad p_z = \sum_{x : x|\Lambda = z} p_x. $$

Thus, for given $\gamma$ and $k$,

$$H(Y(\gamma,\rho)|Z(\gamma,\rho)) = \sum_{y \in \Sigma|\Lambda : x|\Lambda = y} -p_y \log p_y + \sum_{z \in \Sigma|\Lambda : x|\Lambda = z} p_z \log p_z $$

is a function of the probabilities $(p_x)_\gamma$ of $X(M)$.

Our next step will be to reason as follows: We have found linear requirements that the $p_x$’s satisfy and expressed $H(Y(\gamma,\rho)|Z(\gamma,\rho))$ as a function of $(p_x)_x$. However, we do not know of a way to actually calculate $(p_x)_x$. So, instead of the probabilities $(p_x)_x$, consider the variables $(\bar{p}_x)_x$. From this line of thought we get our main theorem.

**Theorem 7:** The value of the optimization program given in Figure 3 is an upper bound on $\text{cap}(S)$.

**Proof:** First, notice that if we take $\bar{p}_x = p_x$, then (by Section IV) all the requirements which the $\bar{p}_x$’s are subject to indeed hold, and the objective function is equal to

$$\sum_{k=1}^\ell \rho(k) c(k) \sum_{\gamma=1}^c \frac{H(Y(\gamma,\rho)|Z(\gamma,\rho))}. $$
maximize \[ \sum_{k=1}^{\ell} p^{(k)} \sum_{\gamma=1}^{c^{(k)}} \Xi(k, \gamma) \]
over the variables \((\vec{p}_x)_{x \in S[|\Lambda|]}\), where for
1 \(\leq k \leq \ell\), 1 \(\leq \gamma \leq c^{(k)}\), \(y \in S[|\Upsilon^{(k)}|]\), \(z \in S[|\Psi^{(k)}|]\),
we define
\[ p^{(k)}_{\gamma,y} \equiv \sum_{x \in S[|\Lambda|]: x|\Upsilon^{(k)}|=y} \vec{p}_x, \quad p^{(k)}_{\gamma,z} \equiv \sum_{x \in S[|\Lambda|]: x|\Psi^{(k)}|=z} \vec{p}_x, \]
\[ \Xi(k, \gamma) \equiv - \sum_{y \in S[|\Upsilon^{(k)}|]} \log p^{(k)}_{\gamma,y} / p^{(k)}_{\gamma,z} + \sum_{z \in S[|\Psi^{(k)}|]} \log p^{(k)}_{\gamma,z} / p^{(k)}_{\gamma,z}, \]
and the variables \(\vec{p}_x\) are subject to the following requirements:
(i) \[ \sum_{x \in S[|\Lambda|]} \vec{p}_x = 1. \]
(ii) For all \(x \in S[|\Lambda|]\), \(\vec{p}_x \geq 0\).
(iii) For all \(x' \in S[|\Lambda'|]\),
\[ \sum_{x : x|\Lambda'|=x'} \vec{p}_{x'} = \sum_{x : x|\sigma_{1,0}(\Lambda')|=\sigma_{1,0}(x')} \vec{p}_x. \]
(iv) For all \(x'' \in S[|\Lambda''|]\),
\[ \sum_{x : x|\Lambda''|=x''} \vec{p}_{x''} = \sum_{x : x|\sigma_{0,1}(\Lambda'')|=\sigma_{0,1}(x'')} \vec{p}_x. \]
(v) (If \(S\) is reflection (resp. complementation) invariant) For all \(x \in S[|\Lambda|]\),
\[ \vec{p}_x = \vec{p}_{n,0}(x) = \vec{p}_{0,0}(x) \quad (\text{resp. } \vec{p}_x = \vec{p}_{\text{comp}}(x)). \]
(vi) (If \(S\) is transposition invariant) For all \(\chi \in S[|\Lambda|]\),
\[ \sum_{x : x|\Lambda|=\chi} \vec{p}_x = \sum_{x : x|\Lambda|=\tau(\chi)} \vec{p}_x. \]

Thus, it suffices to show that each summand is concave in \((\vec{p}_x)_x\). This is indeed the case: let (\(\vec{p}^{(1)}_x\)) \(_{x \in S[|\Lambda|]}\) and (\(\vec{p}^{(2)}_x\)) \(_{x \in S[|\Lambda|]}\) be non-negative. Let 0 \(\leq \xi \leq 1\) be given, and define (\(\vec{p}^{(3)}_x\)) \(_{x \in S[|\Lambda|]}\) as
\[ \vec{p}^{(3)}_x = \xi \vec{p}^{(1)}_x + (1 - \xi) \vec{p}^{(2)}_x, \quad x \in S[|\Lambda|]. \]
For \(t = 1, 2, 3\), denote by \(\vec{p}^{(t)}_y\) and \(\vec{p}^{(t)}_{\gamma,z}\) the marginalizations corresponding to (\(\vec{p}^{(t)}_x\)) \(_x\). Obviously,
\[ \vec{p}^{(3)}_y = \xi \vec{p}^{(1)}_y + (1 - \xi) \vec{p}^{(2)}_y, \quad y \in S[|\Upsilon^{(k)}|], \]
and
\[ \vec{p}^{(3)}_{\gamma,z} = \xi \vec{p}^{(1)}_{\gamma,z} + (1 - \xi) \vec{p}^{(2)}_{\gamma,z}, \quad z \in S[|\Psi^{(k)}|]. \]
We must show that for all \(y \in S[|\Upsilon^{(k)}|], z = y|\Psi^{(k)}]\)
\[ \vec{p}^{(3)}_y \log \frac{\vec{p}^{(3)}_y}{\vec{p}^{(3)}_{\gamma,z}} \leq \xi \vec{p}^{(1)}_y \log \frac{\vec{p}^{(1)}_y}{\vec{p}^{(1)}_{\gamma,z}} + (1 - \xi) \vec{p}^{(2)}_y \log \frac{\vec{p}^{(2)}_y}{\vec{p}^{(2)}_{\gamma,z}}. \]
This is indeed the case, by the log sum inequality \[10\] p. 29.

VI. Computational results
At this point, we have formulated a concave optimization problem, and wish to solve it. There are quite a few programs, termed solvers, that enable one to do so. Many such solvers — most of them proprietary — are hosted on the servers of the NEOS project \[11\][12][13], and the public may submit moderately sized optimization problems to them. We have coded our optimization problems in the AMPL modeling language \[14]\, and submitted them to NEOS.

Essentially, a solver starts with some initial guess as to the minimizing value of (\(\vec{p}_x\)) \(_x\), and then iteratively improves the value of the objective function. This process is terminated when the solver decides that it is “close enough” to the optimum. Denote by \(\overline{\vec{p}} = (\vec{p}_x)_{x \in S[|\Lambda|]}\) this “close enough” assignment to the variables. Of course, we must supply an upper bound on \(|\text{cap}(S)|\), not an approximation to one. Thus, let \(f\) and
\[ \overline{g} = (\overline{g}_x), \quad x \in S[|\Lambda|], \]
be the value of the objective function and its gradient at \(\overline{\vec{p}}\), respectively. Obviously, \(\overline{f}\) is a lower bound on the value of our optimization problem. For an upper bound\[2\] we replace
and get a linear program (the value of which can be calculated exactly). By concavity, the value of this linear program is indeed an upper bound. So, we use NEOS yet again to solve it. For the sake of double-checking, we submitted the above optimization problems to two solvers: IPOPT [15] and MOSEK.

Before stating our computational results, let us first define one more strict total order, which we have termed the “skip” order, \( \prec_{\text{skip}} \) (see Figure 4). We have that \((i_1, j_1) \prec_{\text{skip}} (i_2, j_2)\) iff

- \(i_1 < i_2\), or
- \((i_1 = i_2 \text{ and } j_1 \equiv 0 \pmod{2} \text{ and } j_2 \equiv 1 \pmod{2})\), or
- \((i_1 = i_2 \text{ and } j_1 \equiv j_2 \pmod{2} \text{ and } j_1 < j_2)\)

The first order makes \(\prec_{\text{skip}}\) reflexive; the second makes \(\prec_{\text{skip}}\) asymmetric. When available, these compared-to bounds are listed in Table II. When available, we may define a stationary 1-D Markov chain [1, Theorem 3.17] and its proof, there exists a series of 1-D constraints \(\hat{S}\) such that

\[
\hat{r}, \hat{s}, \hat{t} \equiv 0 \pmod{2}
\]

\[\hat{S} > 0\] and positive integers \(r, s, t\) such that

\[
\hat{S} = \lim_{m \to \infty} \text{cap}(S_m)
\]

To finish the proof, we now show that

\[
\mu(r_0, s_0, t_0) \leq \text{cap}(S_m)
\]

where

\[
r_0 = m + 1, \quad s_0 = 2 \cdot \theta, \quad t_0 = \theta - 1.
\]

Note that \(\mu(r_0, s_0, t_0)\) is the maximum of

\[
H(\tilde{X}_{m, \theta - 1}^{(\leq \epsilon)} \cap B_{m + 1, 2} \theta)];
\]

over all random variables \(\tilde{X} \in S_{m + 1, 2} \theta\) with a probability distribution satisfying our linear requirements.

For all \(0 \leq \phi < \theta\) we get by the (imposed) stationarity of \(\tilde{X}\) that (13) is bounded from above by

\[
H_{\phi} = H(\tilde{X}_{m, \phi}^{(\leq \epsilon)} \cap B_{m + 1, \phi});
\]

so, (13) is also bounded from above by

\[
1 \theta - 1 \theta - 1 \sum_{\phi=0}^{\theta - 1} H_{\phi}.
\]

The first \(\theta\) columns of \(\tilde{X}\) form a configuration with index set \(B_{m + 1, \phi}\). By our linear requirements, stationarity (specifically, vertical stationarity) holds for this configuration as well. So, we may define a stationary 1-D Markov chain [1, §3.2.3] on

\[
\lambda
\]

Fig. 4. An entry labeled \(i\) in the configuration precedes an entry labeled \(j\) according to \(\prec_{\text{skip}}, \text{iff } i < j\).
TABLE I

| Constraint   | r  | s  | k | < used | Upper bound | Comparison | Lower bound |
|--------------|----|----|---|--------|-------------|------------|-------------|
| (2, \infty) RLL | 3 | 7 | 8 | \leq \text{lex} \text{, } \leq \text{skip} | 0.4457 | 0.4459 | 0.16 | 0.444202 |
| (3, \infty) RLL | 4 | 8 | 5 | \leq \text{lex} | 0.36821 | 0.3686 | 0.16 | 0.365623 |
| (0, 2) RLL  | 3 | 2 | 5 | \leq \text{lex} | 0.816731 | 0.817053 | 0.16 | 0.816007 |

\[ S_m, \text{ with entropy given by (14). That entropy, in turn, is at most } \hat{\text{cap}}(S_m). \]

**Proof of Theorem 9**. The following inequalities are easily verified:

\[
\mu(r, s, t) \geq \mu(r + 1, s, t).
\]

\[
\mu(r, s, t) \geq \mu(r, s + 1, t).
\]

\[
\mu(r, s, t) \geq \mu(r, s + 1, t + 1).
\]

The proof follows from them and Lemma 10.

**APPENDIX**

Our goal in this appendix is to prove Theorem 1. Essentially, Theorem 1 will turn out to be a corollary of [3, Theorem 1.4]. However, [3, Theorem 1.4] deals with configurations in which the index set is \( \mathbb{Z}^2 \). So, some definitions and auxiliary lemmas are in order.

Recall that \((G_{\text{row}}, G_{\text{col}})\) is the pair of vertex-labeled graphs through which \( S = S(G_{\text{row}}, G_{\text{col}}) \) is defined. Also, recall that each member of \( S \) is a configuration with a rectangular index set. Namely, the index set of a configuration in \( S \) is \( \sigma_{i,j}(B_{M,N}) \), for some \( i, j, M, \) and \( N \). We now give a very similar definition to that of \( S \), only now we require that the index set of each configuration is \( \mathbb{Z}^2 \). Namely, define \( S = S(G_{\text{row}}, G_{\text{col}}) \) as follows: A configuration \((u_{i,j})(i,j)\in\mathbb{Z}^2\) over \( \Sigma \) is in \( S(G_{\text{row}}, G_{\text{col}}) \) iff there exists a configuration \((u_{i,j})(i,j)\in\mathbb{Z}^2\) over the vertex set \( V \) with the following properties: for all \( (i,j) \in \mathbb{Z}^2 \), (a) the labeling of \( u_{i,j} \) satisfies \( L(u_{i,j}) = u_{i,j} \); (b) there exists an edge from \( u_{i,j} \) to \( u_{i,j+1} \) in \( G_{\text{row}} \); (c) there exists an edge from \( u_{i,j} \) to \( u_{i+1,j} \) in \( G_{\text{col}} \).

For positive integers \( M, N > 0 \), define \( S_{M,N} \) as the restriction of \( S \) to \( B_{M,N} \). Namely,

\[
S_{M,N} = S(B_{M,N}),
\]

where the definition of the restriction operation is as in (3). Also, for \( M \) equal to \( N \), define

\[
S_{M} = S_{M,M}.
\]

Note that for all \( M, N > 0 \) we have

\[
S_{M,N} \subseteq S_{M,N},
\]

and there are cases in which the inclusion is strict. Next, define the capacity of \( S \) as

\[
\text{cap}(S) = \lim_{M \to \infty} \frac{1}{M^2} \log_2 |S_M|.
\]

The limit indeed exists, by sub-additivity (see [3, Appendix], and references therein).

For integers \( M, N > 0 \) and \( \delta \geq 0 \), denote

\[
C_{M,N,\delta} = \sigma_{-\delta,-\delta}(B_{M+2\delta,N+2\delta})
\]

and let

\[
S_{M,N,\delta} = \sigma[C_{M,N,\delta}].
\]

Note that the index set \( C_{M,N,\delta} \) of each element of \( S_{M,N,\delta} \) is simply \( B_{M,N} \), padded with \( \delta \) columns to the right and left and \( \delta \) rows to the top and bottom. The following lemma will help us bridge the gap between finite and infinite index sets.

**Lemma 11**: Let \( w \) be a configuration over the finite alphabet \( \Sigma \) with index set \( B_{M,N} \). If for all \( \delta \geq 0 \) we have that

\[
w \in S_{M,N,\delta},
\]

Then we must have that

\[
w \in S_{M,N}.
\]

**Proof**: Define the following auxiliary directed graph. The vertex set is

\[
\bigcup_{\delta \geq 0} \{ \hat{w} \in S_{M,N,\delta} : \hat{w}|B_{M,N} = w \}.
\]

For every \( \delta \geq 0 \), there is a directed edge from \( w_1 \in S_{M,N,\delta} \) to \( w_2 \in S_{M,N,\delta+1} \) iff \( w_1 = w_2|[C_{M,N,\delta}] \). It is easily seen that this graph is a directed tree with root \( w \), as defined in [19, §2.4]. Since (16) holds for all \( \delta \geq 0 \), the vertex set of the tree is infinite (and countable). On the other hand, since the alphabet size \( |\Sigma| \) is finite, the out-degree of each vertex is finite. Thus, by König’s Infinity Lemma [19, Theorem 2.8], we must have an infinite path in the tree starting from the root \( w \).

Denote the vertices of the above-mentioned infinite path as \( w = w[0], w[1], w[2], \ldots \).

We now show how to find a configuration \((u_{i,j})(i,j)\in\mathbb{Z}^2\) such that \( w' \in S \) and \( w' = w'[B_{M,N}] \). For each \( (i,j) \in \mathbb{Z}^2 \), define \( u'_{i,j} \) as follows: let \( \delta \geq 0 \) be such that \( (i,j) \in C_{M,N,\delta} \), and take \( u'_{i,j} = w[i,j]^{[\delta]} \). It is easily seen that \( w' \) is well defined and contained in \( S \).

The following lemma states that although the inclusion in (15) may be strict, the capacities of \( S \) and \( \hat{S} \) are equal.

**Lemma 12**: Let \( S \) and \( \hat{S} \) be as previously defined. Then,

\[
\text{cap}(S) = \text{cap}(\hat{S}).
\]

**Proof**: By (15), we must have that \( \text{cap}(S) \leq \text{cap}(\hat{S}) \). For the other direction, it suffices to prove that for all \( M > 0 \),

\[
\text{cap}(S) \leq \frac{1}{M^2} \log_2 |S_M|.
\]
So, let us fix $M$ and prove the above. By Lemma [11] there exists $\delta \geq 0$ such that for all $w \in \Sigma^{M \times M}$,

$$w \notin S_M \implies w \notin S_{M,M,\delta}[B_M].$$

For $t > 0$, let $M'$ be shorthand for

$$M' = t \cdot M.$$ 

By the definition of capacity, we have that

$$\text{cap}(S) = \lim_{t \to \infty} \frac{1}{(M')^2} \log_2 |S_{M'}|. \quad (19)$$

Now, let us partition $B_{M'}$ into the following disjoint sub-sets of indexes: for $0 \leq i, j < t$, define the set

$$D_{i,j} = \sigma_{i,M,j,M}(B_M).$$

Let $w' \in S_{M'}$. Notice that for all $0 \leq i, j < t$ for which

$$\sigma_{i,M,j,M}(C_{M,M,\delta}) \subseteq B_{M'}, \quad (20)$$

we must have that $w'[D_{i,j}]$ is equal to some correspondingly shifted element of $S_M$. On the other hand, for $M$ and $\delta$ fixed, the number of pairs $(i, j)$ for which (20) does not hold is $O(t)$. Thus, a simple calculation gives us that

$$\frac{1}{(M')^2} \log_2 |S_{M'}| \leq 1 \frac{M^2}{M^2} \log_2 |S_M| + O(1/t).$$

This, together with (19), proves (18).

For a given $M > 0$, define the set $F(M)$ of configurations with index set $\mathbb{Z}^2$ as follows: a configuration $(w_{i,j})_{(i,j)\in \mathbb{Z}^2}$ is in $F(M)$ iff for all $(i, j) \in \mathbb{Z}^2$,

$$w[\sigma_{i,j}(B_M)] \in S_M.$$ 

Namely, each $M \times M$ “patch” is a correspondingly shifted element of $S_M$.

Note that there exist vertex-labeled graphs $G_{\text{row}}(M)$ and $G_{\text{col}}(M)$ such that $F(M) = S(G_{\text{row}}(M), G_{\text{col}}(M))$. Specifically, the vertex set of both graphs is equal to $S_M$: the label of each such vertex is its lower-left entry; there is an edge from $w_1 \in S_M$ to $w_2 \in S_M$ in $G_{\text{row}}(M)$ ($G_{\text{col}}(M)$) iff the first $M - 1$ rows (columns) of $w_1$ are equal to the last $M - 1$ (rows) columns of $w_2$. Thus, $\text{cap}(F(M))$ exists. Also, since $w \in S$ implies $w \in F(M)$, we have

$$\text{cap}(S) \leq \text{cap}(F(M)). \quad (21)$$

The following is a direct corollary of [8, Theorem 1.4].

**Corollary 13:** For all $M > 0$, there exists a stationary random variable $W(M)$ taking values on $F(M)[B_M]$ such that

$$\text{cap}(F(M)) \leq \frac{1}{M^2} H(W(M)). \quad (22)$$

**Proof of Theorem 7:** Notice that

$$F(M)[B_M] = S_M \subseteq S_M.$$ 

Thus, take $W(M)$ as in Corollary 13 and notice that it satisfies conditions (i) and (ii) in Theorem 1. From (17), (21), and (22) we get that

$$\text{cap}(S) \leq \lim_{M \to \infty} \frac{1}{M^2} \log_2 H(W(M)).$$

But since $W^{(M)}$ takes values on $S_M$, we have by [9, Page 19] that the above inequality is in fact an equality. Thus, condition (iii) is proved.