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, 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 Σ be a finite alphabet. A two-dimensional (2-D) constraint is a set $\mathbb{S}$ of rectangular arrays over Σ. To be called a constraint, $\mathbb{S}$ must satisfy some requirements, formally defined in [1, §1]. Examples of 2-D constraints include the square constraint [2], 2-D runlength-limited (RLL) constraints [3], 2-D symmetric runlength-limited (SRLL) constraints [4], and the "no isolated bits" constraint [5].

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

$$\mathbb{S}_{M,N} = \mathbb{S} \cap \Sigma^{M\times N} , \quad \mathbb{S}_M = \mathbb{S} \cap \Sigma^{M\times M} .$$

The capacity of $\mathbb{S}$ is equal to

$$\mathsf{cap}(\mathbb{S}) = \lim_{M\to\infty} \frac{1}{M^2} \cdot \log_2 |\mathbb{S}_M| . \quad (1)$$

In this paper, we show a method for calculating an upper bound on $\mathsf{cap}(\mathbb{S})$. Other methods for calculating upper bounds on the capacity of 2-D constraints include the stripe method and the method presented by Forchhammer and Justesen [6].

2-D constraints are a generalization of one-dimensional (1-D) constraints (see [7]). In the case of 1-D constraints, it is well-known that the capacity can be expressed as the output 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 [7, §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 $\mathsf{cap}(\mathbb{S})$. 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 (see Table I). Also, our method easily generalizes to higher dimensions.

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 $\mathsf{U} \subseteq \mathbb{Z}^2$ is a set of integer pairs. A 2-D configuration over Σ with an index set U is a function $w : \mathsf{U} \to \Sigma$. We denote such a configuration as $w = (w_{i,j})_{(i,j)\in\mathsf{U}}$, where for all $(i,j) \in \mathsf{U}$, we have that $w_{i,j} \in \Sigma$. In many cases, the index set U will be the $M \times N$ rectangle

$$\mathsf{B}_{M,N} = \{(i,j) : 0 \le i < M , \quad 0 \le j < N\} .$$

Also, denote

$$\mathsf{B}_M = \mathsf{B}_{M,M} = \{(i,j) : 0 \le i,j < M\} .$$

For integers α, β we denote the shifting of U by (α, β) as

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

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

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

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

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

We denote the restriction of $\mathbb{S}$ to U by $\mathbb{S}[\mathsf{U}]$:

$$\mathbb{S}[\mathsf{U}] = \{w : \text{there exists } w' \in \mathbb{S} \text{ such that } w'[\mathsf{U}] = w\} . \quad (2)$$

* This work was supported by grant No. 2002197 from the United-States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

$\prec_{\text{lex}}$

1	2	3	4	5
16	17	18	19	20
6	7	8	9	10
21	22	23	24	25
11	12	13	14	15

$\prec_{\text{irs}}$

Fig. 1. An entry labeled i in the left (right) configuration precedes an entry labeled j according to $\prec_{\text{lex}}$ ($\prec_{\text{irs}}$) iff $i < j$.

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 $\mathsf{T}_{i,j}^{(\prec)}$ as all the indexes preceding (i, j). Namely,

$$\mathsf{T}_{i,j}^{(\prec)} = \left\{ (i', j') \in \mathbb{Z}^2 : (i', j') \prec (i, j) \right\} .$$

In order to enhance the exposition, we give 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.)

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

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

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

We state the following corollary of [8, Theorem 1.4] without proof.

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

$$\mathsf{cap}(\mathbb{S}) = \lim_{M \to \infty} \frac{1}{M^2} \cdot H(W^{(M)}) . \tag{3}$$

We now proceed towards deriving Lemma 2 below, which gives an upper bound on $\mathsf{cap}(\mathbb{S})$, and makes use of the stationarity property. We note in advance that this bound is not actually meant to be calculated, but it will eventually lead to a computable bound in the following sections.

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

$$\Lambda = \mathsf{B}_{r,s} .$$

We will refer to Λ 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 Λ,
- 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 $1 \le \gamma \le c$,

  $$(a_\gamma, b_\gamma) \in \Lambda$$

  (namely, each color γ has a designated point in the patch, *which may or may not be of color* γ).

The function f must satisfy two requirements, which we now elaborate on. Our first requirement is: for all $1 \le \gamma \le c$,

$$\lim_{M \to \infty} \frac{\{(i, j) \in \mathsf{B}_M : f(i, j) = \gamma\}}{M^2} = \frac{1}{c} . \tag{4}$$

Namely, as the orders of $W^{(M)}$ tend to infinity, each color is equally[1] 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 $(i, j) \in \mathbb{Z}^2$,

$$\sigma_{i',j'}(\Psi_\gamma) = \mathsf{T}_{i,j}^{(\prec)} \cap \sigma_{i',j'}(\Lambda) , \tag{5}$$

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 Λ such that (a_γ, b_γ) is shifted to (i, j). Now, consider the set of all indexes in the shifted Λ which precede (i, j): this set must be equal to the correspondingly shifted Ψ_γ.

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:

$$f_{\text{irs}}(i, j) = \begin{cases} 1 & i \equiv 0 \pmod 2 \\ 2 & i \equiv 1 \pmod 2 \end{cases} .$$

Take $(a_1 = 3, b_1 = 5)$ and $(a_2 = 2, b_2 = 4)$. See also Figure 2(b).

[1] In fact, it is possible to generalize (4), and require only that the limit exists for all γ. We have not found this generalization useful.

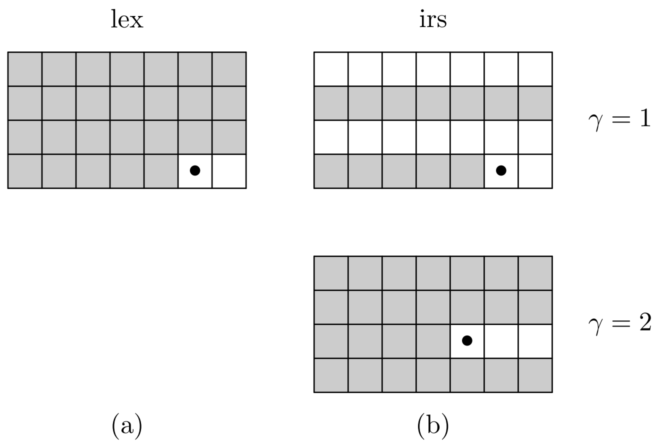

Fig. 2. The left (right) column corresponds to Running Example I (II). The configurations are of order $r \times s$ and represent the index set Λ. The $\bullet$ symbol is in position (a_γ, b_γ). The shaded part is Ψ_γ.

Lemma 2: Let $(W^{(M)})_{M=1}^{\infty}$ be as in Theorem 1 and define

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

Let $\prec$, r, s, c, f, $(\Psi_\gamma)_{\gamma=1}^{c}$, and $(a_\gamma, b_\gamma)_{\gamma=1}^{c}$ be given. For $1 \le \gamma \le c$, define

$$\Upsilon_\gamma = \{(a_\gamma, b_\gamma)\} \cup \Psi_\gamma \; .$$

Let

$$Y_\gamma = X^{(M)}[\Upsilon_\gamma] \;\text{ and }\; Z_\gamma = X^{(M)}[\Psi_\gamma]$$

(note that Y_γ and Z_γ are functions of M). Then,

$$\mathsf{cap}(\mathbb{S}) \le \limsup_{M \to \infty} \frac{1}{c} \sum_{\gamma=1}^{c} H(Y_\gamma | Z_\gamma) \; .$$

Proof: Let X, W and $\mathsf{T}_{i,j}$ be shorthand for $X^{(M)}$, $W^{(M)}$ and $\mathsf{T}_{i,j}^{(\prec)}$, respectively. First note that

$$Y_\gamma = W[\Upsilon_\gamma] \;\text{ and }\; Z_\gamma = W[\Psi_\gamma] \; .$$

We show that

$$\lim_{M \to \infty} \frac{1}{M^2} H(W) \le \limsup_{M \to \infty} \frac{1}{c} \sum_{\gamma=1}^{c} H(Y_\gamma | Z_\gamma) \; .$$

Once this is proved, the claim follows from (3).

By the chain rule [9, Theorem 2.5.1], we have

$$H(W) = \sum_{(i,j) \in \mathsf{B}_M} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \; .$$

We now recall (5) and define the index set $\bar{\partial}$ to be the largest subset of B_M for which the following condition holds: for all $(i,j) \in \bar{\partial}$, we have that

$$\sigma_{i',j'}(\Psi_\gamma) \subseteq \mathsf{B}_M \; , \tag{6}$$

where hereafter in the proof, $\gamma = f(i,j)$, $i' = a_\gamma - i$, and $j' = b_\gamma - j$. Define $\partial = \mathsf{B}_M \setminus \bar{\partial}$. 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

$$\frac{1}{M^2} \sum_{(i,j) \in \partial} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \le \log_2 |\Sigma| \cdot O(1/M) \; .$$

On the other hand, from (5) and (6) we have that for all $(i,j) \in \bar{\partial}$,

$$\sigma_{i',j'}(\Psi_\gamma) \subseteq \mathsf{T}_{i,j} \cap \mathsf{B}_M \; .$$

Hence, since conditioning reduces entropy [9, Theorem 2.6.5],

$$\begin{aligned}
&\frac{1}{M^2} \sum_{(i,j) \in \bar{\partial}} H(W_{i,j} | W[\mathsf{T}_{i,j} \cap \mathsf{B}_M]) \\
&\le \frac{1}{M^2} \sum_{(i,j) \in \bar{\partial}} H(W_{i,j} | W[\sigma_{i',j'}(\Psi_\gamma)]) \\
&= \frac{1}{M^2} \sum_{(i,j) \in \bar{\partial}} H(W[\{(i,j)\} \cup \sigma_{i',j'}(\Psi_\gamma)] | W[\sigma_{i',j'}(\Psi_\gamma)]) \\
&= \frac{1}{M^2} \sum_{(i,j) \in \bar{\partial}} H(Y_\gamma | Z_\gamma) \; ,
\end{aligned}$$

where the last step follows from the stationarity of $W^{(M)}$. Recalling (4), 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 ℓ be a positive integer, and let $(\rho^{\langle k \rangle})_{k=1}^{\ell}$ be non-negative reals such that $\sum_{k=1}^{\ell} \rho^{\langle k \rangle} = 1$. For every $1 \le k \le \ell$, let $\prec^{\langle k \rangle}$, $c^{\langle k \rangle}$, $f^{\langle k \rangle}$, $(\Psi_\gamma^{\langle k \rangle})_{\gamma=1}^{c}$, and $(a_\gamma^{\langle k \rangle}, b_\gamma^{\langle k \rangle})_{\gamma=1}^{c}$ be given. Also, for $1 \le \gamma \le c^{\langle k \rangle}$, let

$$\Upsilon_\gamma^{\langle k \rangle} = \{(a_\gamma^{\langle k \rangle}, b_\gamma^{\langle k \rangle})\} \cup \Psi_\gamma^{\langle k \rangle} \; .$$

Define

$$Y_\gamma^{\langle k \rangle} = X^{(M)}[\Upsilon_\gamma^{\langle k \rangle}] \;\text{ and }\; Z_\gamma^{\langle k \rangle} = X^{(M)}[\Psi_\gamma^{\langle k \rangle}]$$

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

$$\mathsf{cap}(\mathbb{S}) \le \limsup_{M \to \infty} \sum_{k=1}^{\ell} \frac{\rho^{\langle k \rangle}}{c^{\langle k \rangle}} \sum_{\gamma=1}^{c^{\langle k \rangle}} H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle}) \; .$$

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

$$\begin{gathered}
\ell = 1 \; , \quad \rho^{\langle 1 \rangle} = 1 \; , \quad c^{\langle 1 \rangle} = 1 \; , \quad \prec^{\langle 1 \rangle} = \prec_{\text{lex}} \; , \\
(a_1^{\langle 1 \rangle}, b_1^{\langle 1 \rangle}) = (r-1, t) \; , \quad \text{and} \quad \Psi_1^{\langle 1 \rangle} = \Lambda \cap \mathsf{T}_{(a_1^{\langle 1 \rangle}, b_1^{\langle 1 \rangle})} \; ,
\end{gathered} \tag{7}$$

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

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 $x \in \mathbb{S}[\Lambda]$ be a realization of X. Denote

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

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

$$p_x \geq 0 \; .$$

Also,

$$\sum_{x \in \mathbb{S}[\Lambda]} p_x = 1 \; .$$

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

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

For $x' \in \mathbb{S}[\Lambda']$ we must have by stationarity that

$$\mathrm{Prob}(X[\Lambda'] = x') = \mathrm{Prob}(X[\sigma_{1,0}(\Lambda')] = \sigma_{1,0}(x')) \; . \quad (8)$$

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

$$\mathrm{Prob}\left(X = \begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ * & * & * \end{smallmatrix} \right) = \mathrm{Prob}\left(X = \begin{smallmatrix} * & * & * \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right) \; ,$$

where $*$ denotes "don't care".

Both the left-hand and right-hand sides of (8) are marginalizations of $(p_x)_x$. Thus, we get a set of linear equations on $(p_x)_x$, namely, for all $x' \in \mathbb{S}[\Lambda']$,

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

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

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

for all $x'' \in \mathbb{S}[\Lambda'']$,

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

B. Linear equations from reflection, transposition, and complementation

We now show that if $\mathbb{S}$ 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 τ as the transposition of a configuration. Namely,

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

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

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

We say that $\mathbb{S}$ is reflection invariant if for all $M > 0$ and $w \in \Sigma^{M \times M}$,

$$w \in \mathbb{S} \iff h_M(w) \in \mathbb{S} \iff v_M(w) \in \mathbb{S} \; .$$

Transposition and complementation invariance is defined similarly.

The following three claims are easily proved. (a) Suppose that $\mathbb{S}$ is reflection invariant. Then w.l.o.g., for all $x \in \mathbb{S}[\Lambda]$,

$$p_x = p_{v_r(x)} = p_{h_s(x)} \; .$$

(b) Suppose that $\mathbb{S}$ is transposition invariant. Also, assume w.l.o.g., that $r \leq s$, and let

$$\tilde{\Lambda} = \{(i,j) : 0 \leq i,j < r\} \; .$$

Then, w.l.o.g., for all $\chi \in \mathbb{S}[\tilde{\Lambda}]$,

$$\sum_{x \, : \, x[\tilde{\Lambda}]=\chi} p_x = \sum_{x \, : \, x[\tilde{\Lambda}]=\tau(\chi)} p_x \; .$$

(c) Suppose that $\mathbb{S}$ is reflection invariant. Then w.l.o.g., for all $x \in \mathbb{S}[\Lambda]$,

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

V. An upper bound on $\mathsf{cap}(\mathbb{S})$

For the rest of this section, let r, s, ℓ, $\rho^{\langle k \rangle}$, $\prec^{\langle k \rangle}$, $c^{\langle k \rangle}$, $f^{\langle k \rangle}$, $\Psi_\gamma^{\langle k \rangle}$, and $(a_\gamma^{\langle k \rangle}, b_\gamma^{\langle k \rangle})$ be given as in Corollary 3. Recall from Corollary 3 that we are interested in $H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle})$, in order to bound $\mathsf{cap}(\mathbb{S})$ from above.

As a first step, we fix M and express $H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle})$ in terms of the probabilities $(p_x)_x$ of the random variable $X^{(M)}$. For given $1 \leq k \leq \ell$ and $1 \leq \gamma \leq c^{\langle k \rangle}$, let

$$y \in \mathbb{S}[\Upsilon_\gamma^{\langle k \rangle}] \;\; \text{and} \;\; z \in \mathbb{S}[\Psi_\gamma^{\langle k \rangle}]$$

be realizations of $Y_\gamma^{\langle k \rangle}$ and $Z_\gamma^{\langle k \rangle}$, respectively. Let

$$p_{\gamma,y}^{\langle k \rangle} = \mathrm{Prob}(Y_\gamma^{\langle k \rangle} = y) \quad \text{and} \quad p_{\gamma,z}^{\langle k \rangle} = \mathrm{Prob}(Z_\gamma^{\langle k \rangle} = z)$$

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

$$p_y = \sum_{x \in \mathbb{S}[\Lambda] \, : \, x[\Upsilon_\gamma^{\langle k \rangle}]=y} p_x \; , \quad p_z = \sum_{x \in \mathbb{S}[\Lambda] \, : \, x[\Psi_\gamma^{\langle k \rangle}]=z} p_x \; .$$

Thus, for given γ and k,

$$H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle}) = \sum_{y \in \mathbb{S}[\Upsilon_\gamma^{\langle k \rangle}]} -p_y \log_2 p_y + \sum_{z \in \mathbb{S}[\Psi_\gamma^{\langle k \rangle}]} p_z \log_2 p_z$$

is a function of the probabilities $(p_x)_x$ 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 $H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle})$ as a function of $(p_x)_x$. However, *we do not*

TABLE I

UPPER-BOUNDS ON THE CAPACITY OF SOME 2-D CONSTRAINTS.

Constraint	r	s	k	Upper bound	Comparison
$(2,\infty)$-RLL	3	8	7	0.4457	0.4459 [11]
$(3,\infty)$-RLL	4	8	5	0.36821	0.3686 [11]
$(0,2)$-RLL	3	5	2	0.816731	0.817053
n.i.b.	3	4	1	0.92472	0.927855

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 4: The value of the optimization program given in Figure 3 is an upper bound on $\mathsf{cap}(\mathbb{S})$.

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

$$\sum_{k=1}^{\ell} \frac{\rho^{\langle k \rangle}}{c^{\langle k \rangle}} \sum_{\gamma=1}^{c^{\langle k \rangle}} H(Y_\gamma^{\langle k \rangle} | Z_\gamma^{\langle k \rangle}) \ .$$

So, the maximum is an upper bound on the above equation. Next, by compactness, a maximum indeed exists. Since the maximum is not a function of M, the claim now follows from Corollary 3. ■

We must now show that the optimization problem in Figure 3 is an instance of concave programming [10, p. 137], and thus easily calculated. Since the requirements that the variables $(\bar{p}_x)_x$ are subject to are linear, this reduces to showing that the objective function is concave in $(\bar{p}_x)_x$. This is indeed the case; the proof essentially follows from the log sum inequality [9, p. 29] and is omitted due to space limitations.

Our computational results appear in Table I. To the best of our knowledge, they are presently the tightest. We compare our results to those obtained by the method described in [6]. When available, these compared-to bounds are taken from previously published work (specifically, ref. [11].) The rest are the result of our implementation of [6]. We note that the lexicographic orders used to obtain our results were $\prec_{\mathrm{lex}}$, and a strict total order which we denote by $\prec_{\mathrm{skip}}$, and is defined as follows: $(i_1, j_1) \prec_{\mathrm{skip}} (i_2, j_2)$ iff

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

maximize

$$\sum_{k=1}^{\ell} \frac{\rho^{\langle k \rangle}}{c^{\langle k \rangle}} \sum_{\gamma=1}^{c^{\langle k \rangle}} \Xi(k, \gamma)$$

over the variables $(\bar{p}_x)_{x \in \mathbb{S}[\Lambda]}$, where for

$$1 \le k \le \ell \ , \quad 1 \le \gamma \le c^{\langle k \rangle} \ , \quad y \in \mathbb{S}[\Upsilon_\gamma^{\langle k \rangle}] \ , \quad z \in \mathbb{S}[\Psi_\gamma^{\langle k \rangle}] \ ,$$

we define

$$\bar{p}_{\gamma,y}^{\langle k \rangle} \triangleq \sum_{x \in \mathbb{S}[\Lambda] \,:\, x[\Upsilon_\gamma^{\langle k \rangle}] = y} \bar{p}_x \ , \quad \bar{p}_{\gamma,z}^{\langle k \rangle} \triangleq \sum_{x \in \mathbb{S}[\Lambda] \,:\, x[\Psi_\gamma^{\langle k \rangle}] = z} \bar{p}_x \ ,$$

$$\Xi(k,\gamma) \triangleq -\sum_{y \in \mathbb{S}[\Upsilon_\gamma^{\langle k \rangle}]} \bar{p}_{\gamma,y}^{\langle k \rangle} \log_2 \bar{p}_{\gamma,y}^{\langle k \rangle} + \sum_{z \in \mathbb{S}[\Psi_\gamma^{\langle k \rangle}]} \bar{p}_{\gamma,z}^{\langle k \rangle} \log_2 \bar{p}_{\gamma,z}^{\langle k \rangle} \ ,$$

and the variables $\bar{p}_x$ are subject to the following requirements:

(i)
$$\sum_{x \in \mathbb{S}[\Lambda]} \bar{p}_x = 1 \ .$$

(ii) For all $x \in \mathbb{S}[\Lambda]$,
$$\bar{p}_x \ge 0 \ .$$

(iii) For all $x' \in \mathbb{S}[\Lambda']$,
$$\sum_{x \,:\, x[\Lambda'] = x'} \bar{p}_x = \sum_{x \,:\, x[\sigma_{1,0}(\Lambda')] = \sigma_{1,0}(x')} \bar{p}_x \ .$$

(iv) For all $x'' \in \mathbb{S}[\Lambda'']$,
$$\sum_{x \,:\, x[\Lambda''] = x''} \bar{p}_x = \sum_{x \,:\, x[\sigma_{0,1}(\Lambda'')] = \sigma_{0,1}(x'')} \bar{p}_x \ .$$

(v) (If $\mathbb{S}$ is reflection (resp. complementation) invariant) For all $x \in \mathbb{S}[\Lambda]$,
$$\bar{p}_x = \bar{p}_{v_r(x)} = \bar{p}_{h_s(x)} \quad (\text{resp. } \bar{p}_x = \bar{p}_{\mathrm{comp}(x)}) \ .$$

(vi) (If $\mathbb{S}$ is transposition invariant) For all $\chi \in \mathbb{S}[\tilde{\Lambda}]$,
$$\sum_{x \,:\, x[\tilde{\Lambda}] = \chi} \bar{p}_x = \sum_{x \,:\, x[\tilde{\Lambda}] = \tau(\chi)} \bar{p}_x \ .$$

Fig. 3. Optimization program over the variables $\bar{p}_x$ (assuming w.l.o.g. that $r \le s$). The optimum is an upper bound on $\mathsf{cap}(\mathbb{S})$.

REFERENCES

[1] S. Halevy and R. M. Roth, "Parallel constrained coding with application to two-dimensional constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1009–1020, 2002.

[2] W. Weeks and R. E. Blahut, "The capacity and coding gain of certain checkerboard codes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1193–1203, 1998.

[3] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained code," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1527–1540, 1999.

[4] T. Etzion, "Cascading methods for runlength-limited arrays," *IEEE Trans. Inform. Theory*, vol. 43, pp. 319–324, 1997.

[5] S. Halevy, J. Chen, R. M. Roth, P. H. Siegel, and J. K. Wolf, "Improved bit-stuffing bounds on two-dimensional constraints," *IEEE Trans. Inform. Theory*, vol. 50, pp. 824–838, 2004.

[6] S. Forchhammer and J. Justesen, "Bounds on the capacity of constrained two-dimensional codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2659–2666, 2000.

[7] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. Pless and W. Huffman, Eds. Amsterdam: Elsevier, 1998, pp. 1635–1764.

[8] R. Burton and J. E. Steif, "Non-uniqueness of measures of maximal entropy for subshifts of finite type," *Ergod. Th. Dynam. Sys.*, vol. 14, pp. 213–235, 1994.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[11] S. Forchhammer and T. V. Laursen, "Entropy of bit-stuffing-induced measures for two-dimensional checkerboard constraints," *IEEE Trans. Inform. Theory*, vol. 53, pp. 1537–1546, 2007.