Multidimensional Semiconstrained Systems

Ohad Elishco
Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
ohadeli@bgu.ac.il

Tom Meyerovitch
Department of Mathematics
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
mtom@math.bgu.ac.il

Moshe Schwartz
Electrical and Computer Engineering
Ben-Gurion University of the Negev
Beer Sheva 8410501, Israel
schwartz@ee.bgu.ac.il

***Abstract*—We generalize the notion of independence entropy to the study of semiconstrained systems. Using it, we obtain a new lower bound on the capacity of multi-dimensional semiconstrained systems. We show the new bound improves upon the best-known bound in a case study of $(0,k,p)$-RLL semiconstrained systems.**

I. Introduction

Error-correcting codes and constrained codes may be considered as two extreme ways of coping with a noisy channel. The former are usually data independent, and assume errors are a statistical phenomenon, reducing data-transmission rate to protect against such errors. Constrained codes, however, assume certain patterns in the data stream are responsible for the occurrence of errors. Thus, constrained codes eliminate all undesirable patterns, at the cost of reduced data-transmission rate.

Recently in [5], [6], semiconstrained systems (SCS) were suggested as a generalization to constrained systems (which we emphasize by calling fully constrained systems). In SCS we do not eliminate the undesirable patterns entirely but rather we allow them to appear with a restriction on their frequency. Informally, a SCS is defined by a set Γ of probability measures over k-tuples. The allowed words in the SCS are those in which the empirical distribution of k-tuples belongs to Γ. This may be viewed as a generalization of fully constrained systems since taking Γ to be a subset with a 0-frequency restriction on some k-tuples yields a fully constrained system.

SCS not only generalize fully constrained systems, but also subsume a range of other settings, which were mainly dealt with in an ad-hoc fashion. Among these we can find DC-free RLL coding [14], constant-weight ICI coding for flash memories [3], [4], [12], [23], coding to mitigate the appearance of ghost pulses in optical communication [26], [27], and the more general, channel with cost constraints [10], [13].

The capacity of SCS is known via large-deviation tools, e.g., [18]. A probabilistic encoder for SCS was constructed in [6], and constant-bit-rate to constant-bit-rate encoders are possible by approximating a SCS with a fully constrained system, as described in [5].

A natural extension, and the goal of this work, is to study multi-dimensional SCS. This is an extremely challenging problem, considering the fact that even for fully constrained system in complete generality it is provably impossible to find an exact solution. The capacity of multi-dimensional fully constrained systems is known exactly only in a handful of cases [1], [15], [17], [24]. In the absence of a general method for computing the capacity, various bounds and approximations were studied, e.g., [2], [8], [9], [11], [20], [21], [25], [28]–[30]. It should be emphasized that apart from its independent intellectual merit, studying multi-dimensional systems is of practical importance since most storage media are two- or three-dimensional, including magnetic recording devices such as hard drives, optical recording devices such as CDs and DVDs, and flash.

The approach we take in this work is bounding the capacity by studying the independence entropy of SCS, thus extending the works [16], [19]. Although the notion of independence entropy was first defined in [16], the idea stemmed from tradeoff functions studied in [22].

The main contributions of this paper are a formulation of the independence entropy for SCS, and a study of the independence entropy and its relation to the capacity of SCS. As a result, we obtain a new lower bound on the capacity of multi-dimensional SCS, improving upon the best known bounds given in [6].

This work is a step towards establishing an equality of limiting entropy and independence entropy. As the independence entropy is a lower bound on the entropy of a given SCS in every dimension, we believe that as the dimension grows, the capacity approaches the independence entropy as in the case of fully constrained systems [19].

This paper is organized as follows. In Section II we describe the notation and give the required definitions used throughout the paper. In Section III we provide results characterizing the independence entropy, and its relation with the capacity of multi-dimensional SCS. We conclude in Section IV by describing a short case study, and compare it with previous results.

II. Preliminaries

Throughout the paper, $\mathbb{N}$ denotes the set of natural numbers, including 0. Given a finite alphabet Σ, we denote by Σ^* the set of all finite words over Σ. For a word $w \in \Sigma^*$, we denote by $|w|$ the length of w. Additionally, for a word $w = w_0 \dots w_{n-1} \in \Sigma^n$, $w_i \in \Sigma$, we define $\mathrm{fr}_w^k : \Sigma^k \to [0,1]$ as the frequency of k-tuples in w, i.e., for any $a \in \Sigma^k$,

$$\mathrm{fr}_w^k(a) \triangleq \frac{1}{|w|} \sum_{i=0}^{|w|-1} \mathbb{1}_a(w_i \dots w_{i+k-1})$$

where all coordinate indices are taken modulo $|w|$, and the function $\mathbb{1}$ is the indicator function. Hence, $\mathrm{fr}_w^k(a)$ is the

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

empirical frequency of the k-tuple a in the word w. By abuse of notation we shall sometimes refer to fr_w^k as a vector, i.e., $\mathrm{fr}_w^k = (\mathrm{fr}_w^k(a_0), \ldots, \mathrm{fr}_w^k(a_{|\Sigma|^k-1}))$ where a_i ranges over Σ^k in some fixed order.

For a set A we denote by $\mathcal{P}(A)$ the set of all probability measures over A. We now define a one-dimensional SCS.

Definition 1. *Let $k \in \mathbb{N}$, $k \geqslant 1$. A* semiconstrained system *(SCS) is defined by a set $\Gamma \subseteq \mathcal{P}(\Sigma^k)$. The language of a SCS Γ is the set*

$$\mathcal{B}(\Gamma) \triangleq \left\{ w \in \Sigma^* \ : \ \mathrm{fr}_w^k \in \Gamma \right\}.$$

We define $\mathcal{B}_n(\Gamma) \triangleq \mathcal{B}(\Gamma) \cap \Sigma^n$.

Definition 2. *The capacity of a given SCS, Γ, is defined as*

$$\mathsf{cap}(\Gamma) \triangleq \limsup_{n\to\infty} \frac{1}{n} \log_2 \left(|\mathcal{B}_n(\Gamma)| \right).$$

The limit in the capacity definition does not necessarily exist but under mild assumptions it does [5]. We introduce these assumptions. First, we need the notion of shift-invariant measures.

Definition 3. *Let $\mu \in \mathcal{P}(\Sigma^k)$. We say that μ is* shift invariant *if*

$$\sum_{b\in\Sigma} \mu(b, a_1, \ldots, a_{k-1}) = \sum_{b\in\Sigma} \mu(a_1, \ldots, a_{k-1}, b),$$

for all $a_1, \ldots, a_{k-1} \in \Sigma$. We denote by $\mathcal{P}_{\mathrm{si}}(\Sigma^k) \subseteq \mathcal{P}(\Sigma^k)$ the set of all shift-invariant probability measures over Σ^k.

For a set $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ we denote by $\mathrm{int}(\Gamma)$ and by $\mathrm{cl}(\Gamma)$ the interior and closure of Γ, respectively.

Definition 4. *For a set $\Gamma \subseteq \mathcal{P}(\Sigma^k)$, we say that Γ is* fat *if*

$$\mathrm{cl}\left(\mathrm{int}\left(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k)\right)\right) = \mathrm{cl}\left(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k)\right).$$

We have the following theorem from [5].

Theorem 5. *[5, Section 2] Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be closed and convex. If Γ is fat then*

$$\mathsf{cap}(\Gamma) = \log_2 |\Sigma| - \inf_{\eta \in \Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k)} H(\eta|\mu)$$

where $H(\cdot|\cdot)$ is the relative entropy function and the measure μ defined by $\mu(\phi a) = \frac{1}{|\Sigma|} \sum_{a' \in \Sigma} \eta(\phi a')$ for all $\phi \in \Sigma^{k-1}$ and $a \in \Sigma$.

Note that if Γ is fat then there are no 0-constraints over k-tuples, otherwise the interior of Γ is empty. Some of our results can be extended to deal with combinatorial constraints as well by defining relatively fat SCS or weak SCS [5, Section 4].

We use $\mathbf{e}_i$ to denote the unit vector of direction i, $\mathbf{0}$ to denote the all-zero vector, and $\mathbf{1}$ to the denote the all-one vector. For $n \in \mathbb{N}$ we define $[n] \triangleq \{0, 1, \ldots, n-1\}$. We shall often use $[n]\mathbf{e}_i$ to denote the set $\{0 \cdot \mathbf{e}_i, 1 \cdot \mathbf{e}_i, \ldots, (n-1) \cdot \mathbf{e}_i\}$. Denote by F_n^d the d-dimensional cube of length n, i.e., the set $F_n^d \triangleq [n]^d$. Additionally, for $(n_0, \ldots, n_{d-1}) \in \mathbb{N}^d$ we conveniently denote $[(n_0, \ldots, n_{d-1})] \triangleq [n_0] \times [n_1] \times \cdots \times [n_{d-1}]$.

For $\mathbf{u} \in F_n^d$ denote by $\sigma_{\mathbf{u}} : \Sigma^{F_n^d} \to \Sigma^{F_n^d}$ the cyclic shift operator, i.e., for a word $w \in \Sigma^{F_n^d}$ and for $\mathbf{v}, \mathbf{u} \in F_n^d$,

$$(\sigma_{\mathbf{u}}(w))_{\mathbf{v}} \triangleq w_{\mathbf{u}+\mathbf{v}}$$

where the coordinates are taken modulo n.

For a word $w \in \Sigma^{F_n^d}$ and for $S \subseteq F_n^d$ we denote by $w_S \in \Sigma^S$ the reduction of w to the coordinates in S. If $x \in \Sigma^S$, define the word indicator function $\mathbb{1}_x : \Sigma^{F_n^d} \to \{0, 1\}$ by $\mathbb{1}_x(w) = 1$ if $w_S = x$ and 0 otherwise. We then denote by $\mathrm{fr}_w^S \in \mathcal{P}(\Sigma^S)$ the frequency of S-tuples in w, i.e., for $x \in \Sigma^S$,

$$\mathrm{fr}_w^S(x) \triangleq \frac{1}{n^d} \sum_{\mathbf{u} \in F_n^d} \mathbb{1}_x(\sigma_{\mathbf{u}}(w)),$$

where the coordinates are taken modulo n.

Let $\mu \in \mathcal{P}(\Sigma^{F_n^d})$ be a measure. For $\mathbf{u} \in F_n^d$ and for $S \subseteq F_n^d$, we denote by $\delta_{\mathbf{u}}^S(\mu)$ the marginal distribution of the coordinates $\{\mathbf{u} + \mathbf{v} \ : \ \mathbf{v} \in S\}$ where the coordinates are taken modulo n, i.e., for $x \in \Sigma^S$, $\delta_{\mathbf{u}}^S(\mu) \in \mathcal{P}(\Sigma^S)$ is defined by

$$\delta_{\mathbf{u}}^S(\mu)(x) \triangleq \sum_{w \in \Sigma^{F_n^d},\ (\sigma_{\mathbf{u}}(w))_S = x} \mu(w).$$

It shall be convenient for us to define the average,

$$\delta^S(\mu) \triangleq \frac{1}{n^d} \sum_{\mathbf{u} \in F_n^d} \delta_{\mathbf{u}}^S(\mu).$$

We are now ready to define the d-axial-product SCS.

Definition 6. *For $i \in [d]$, let $\Gamma_i \subseteq \mathcal{P}(\Sigma^{k_i})$ be one dimensional SCS. Denote $k \triangleq \max_{i\in[d]} k_i$. The d-axial-product SCS, denoted by $\otimes_{i=0}^{d-1}\Gamma_i$ is defined by*

$$\otimes_{i=0}^{d-1}\Gamma_i \triangleq \left\{ \mu \in \mathcal{P}(\Sigma^{F_k^d}) \ : \ \forall i \in [d],\ \delta_{\mathbf{0}}^{[k_i]\mathbf{e}_i} \in \Gamma_i \right\}.$$

Furthermore, for every $n \in \mathbb{N}$ we define

$$\mathcal{B}_n\left(\otimes_{i=0}^{d-1}\Gamma_i\right) \triangleq \left\{ w \in F_n^d \ : \ \forall i \in [d],\ \mathrm{fr}_w^{[k_i]\mathbf{e}_i} \in \Gamma_i \right\},$$

with coordinates taken modulo n. The language of the d-axial-product SCS $\otimes_{i=0}^{d-1}\Gamma_i$ is the set

$$\mathcal{B}(\otimes_{i=0}^{d-1}\Gamma_i) \triangleq \bigcup_{n\in\mathbb{N}} \mathcal{B}_n\left(\otimes_{i=0}^{d-1}\Gamma_i\right).$$

Intuitively, the arrays of a d-axial-product SCS satisfy that along the ith direction, the empirical frequencies of k_i-tuples is in Γ_i. Note that $\otimes_{i=0}^{d-1}\Gamma_i$ induces a set of measures over $\Sigma^{F_k^d}$ where $k = \max_i k_i$., i.e., $\otimes_{i=0}^{d-1}\Gamma_i \subseteq \mathcal{P}(\Sigma^{F_k^d})$. For a set $S \subseteq F_n^d$ with σ-algebra $\mathcal{B}$, consider the space $\mathcal{P}(\Sigma^{F_n^d})$ with the total-variation distance, i.e., the measurable space $(\Sigma^{F_n^d}, \mathcal{B})$ and $\mu, \nu \in \mathcal{P}(\Sigma^{F_n^d})$, with

$$\|\mu - \nu\|_{TV} \triangleq \sup_{A\in\mathcal{B}} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{w \in \Sigma^{F_n^d}} |\mu(w) - \nu(w)|.$$

For any set of measures $M \subseteq \mathcal{P}(\Sigma^{F_n^d})$ and for $\epsilon > 0$, we denote by $\mathbb{B}_\epsilon(M)$ the ϵ-neighborhood of M,

$$\mathbb{B}_\epsilon(M) \triangleq \left\{ \mu \in \mathcal{P}(\Sigma^{F_n^d}) \ : \ \inf_{\nu\in M} \|\mu - \nu\|_{TV} \leqslant \epsilon \right\}.$$

Definition 7. *Let $\otimes_{i=0}^{d-1}\Gamma_i$ be a d-axial-product SCS. Its capacity is defined by,*

$$\mathsf{cap}(\otimes_{i=0}^{d-1}\Gamma_i) \triangleq \limsup_{n\to\infty} \frac{1}{n^d}\log_2 \left|\mathcal{B}_n(\otimes_{i=0}^{d-1}\Gamma_i)\right|.$$

The external capacity is defined as

$$\overline{\mathsf{cap}}(\otimes_{i=0}^{d-1}\Gamma_i) \triangleq \lim_{\epsilon\to 0}\limsup_{n\to\infty} \frac{1}{n^d}\log_2 \left|\mathcal{B}_n\left(\mathbb{B}_\epsilon(\otimes_{i=0}^{d-1}\Gamma_i)\right)\right|.$$

The external capacity is equivalent to the expression $\lim_{\epsilon\to 0}\mathsf{cap}(\mathbb{B}_\epsilon(\otimes_{i=0}^{d-1}\Gamma_i))$. We may think of the external capacity as a relaxation of the constraints given in Γ. It is equivalent to the weak SCS which is considered in [6]. The reason behind a further relaxation of the constraints is avoiding pathological examples. A simple example in the one dimensional case is when we consider all words in which the empirical frequency of the pattern 11 is exactly r when r is irrational. In this case the capacity is $-\infty$ since there are no finite words which satisfy the constraints. Considering the external capacity we obtain a positive value which corresponds to the amount of words in which the empirical distribution of the pattern 11 is in $[r-\epsilon, r+\epsilon]$ when ϵ decays to zero as the length of the word grows.

Let μ_n denote the uniform distribution over $\Sigma^{F_n^d}$ and denote by $\otimes_{i=0}^{d-1}\Gamma_i$ the d-axial-product SCS. We can write

$$\mathsf{cap}(\otimes_{i=0}^{d-1}\Gamma_i) = \log_2|\Sigma| + \limsup_{n\to\infty}\frac{1}{n^d}\log_2 \mu_n\left(\mathcal{B}_n(\otimes_{i=0}^{d-1}\Gamma_i)\right).$$

As in the one-dimensional case, we have large-deviations results for $\mu_n\left(\left\{w \in \mathcal{B}_n(\otimes_{i=0}^{d-1}\Gamma_i)\right\}\right)$ (see, for example, [7]).

Although we have large deviations results for d-dimensional SCSs, the expressions involved are much more complicated than the relative entropy function. Unfortunately, the fat condition does not guarantee that the inequalities are in fact equalities. Thus, there is a necessity for more easily computable bounds on the capacity. To this end, we define the independence entropy of a d-dimensional SCS, which is the basis of the main results of this paper.

III. Independence Entropy

In this section we define the independence entropy of multi-dimensional SCS, and relate it to the capacity. Thus, we obtain a new lower bound on the capacity of multi-dimensional SCS.

Definition 8. *Let $\mu \in \mathcal{P}(\Sigma^S)$. We say that μ is* independent *if $\mu \in (\mathcal{P}(\Sigma))^S$, i.e., for every $S' \subseteq S$, $\delta_{\mathbf{0}}^{S'}(\mu) = \prod_{\mathbf{v}\in S'}\delta_{\mathbf{v}}^{[\mathbf{1}]}(\mu)$.*

In other words, we say that μ is independent if there exists $\{p_{\mathbf{u}} \in \mathcal{P}(\Sigma) \ : \ \mathbf{u} \in S\}$ such that $\mu = \prod_{\mathbf{u}\in S} p_{\mathbf{u}}$.

Let $\Gamma = \otimes_{i=0}^{d-1}\Gamma_i$ be the d-axial product of $\Gamma_i \subseteq \mathcal{P}(\Sigma^{k_i})$. Define

$$\overline{\mathcal{P}}_n(\Gamma) \triangleq \left\{\mu \in (\mathcal{P}(\Sigma))^{F_n^d} \ : \ \forall i \in [d],\ \delta^{[k_i]\mathbf{e}_i}(\mu) \in \Gamma_i\right\},$$

where the coordinates are taken modulo n. Thus, $\overline{\mathcal{P}}_n(\Gamma)$ is the set of all independent probability measures on $\Sigma^{F_n^d}$ such that the average of the $[k_i]\mathbf{e}_i$-marginals is in Γ_i for each $i \in [d]$. We note that if Γ is convex, then $\overline{\mathcal{P}}_n(\Gamma)$ is also convex.

We can now define the independence entropy of a SCS.

Definition 9. *Let $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$ be a d-dimensional SCS. The* independence entropy *of Γ is defined by*

$$h_{\mathrm{ind}}(\Gamma) \triangleq \limsup_{n\to\infty}\sup_{\mu\in\overline{\mathcal{P}}_n(\Gamma)}\frac{1}{n^d}H(\mu),$$

where $H(\mu) = -\sum_{w\in\Sigma^{F_n^d}}\mu(w)\log_2\mu(w)$ is the entropy of μ. We define the external independence entropy by

$$\overline{h_{\mathrm{ind}}}(\Gamma) \triangleq \lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{\mu\in\overline{\mathcal{P}}_n(\mathbb{B}_\epsilon(\Gamma))}\frac{1}{n^d}H(\mu)$$

Before stating the main results of this paper, we focus a bit on $\delta^S(\mu)$. Note that $\delta^S : \mathcal{P}(\Sigma^{F_n^d}) \to \mathcal{P}(\Sigma^S)$.

Definition 10. *A function $f : X \to Y$ between two metric spaces X, Y is called a* contraction *if $d_Y(f(x), f(x')) \leqslant d_X(x, x')$ for every $x, x' \in X$.*

Lemma 11. *Let $S \subseteq F_n^d$ and $\delta_{\mathbf{u}}^S, \delta^S : \mathcal{P}(\Sigma^{F_n^d}) \to \mathcal{P}(\Sigma^S)$ be as defined above. Then $\delta_{\mathbf{u}}^S$ and δ^S are contractions with respect to the total-variation distance, i.e., for all $\mu, \nu \in \mathcal{P}(\Sigma^{F_n^d})$, and all $\mathbf{u} \in F_n^d$,*

$$\|\delta_{\mathbf{u}}^S(\mu) - \delta_{\mathbf{u}}^S(\nu)\|_{TV} \leqslant \|\mu-\nu\|_{TV},$$
$$\|\delta^S(\mu) - \delta^S(\nu)\|_{TV} \leqslant \|\mu-\nu\|_{TV}.$$

Lemma 12. *Let $w \in \Sigma^{F_n^d}$, $k \leqslant n$, and $S \subseteq F_k^d$. Then,*

$$\delta_{\mathbf{0}}^S(\mathrm{fr}_w^{F_k^d}) = \mathrm{fr}_w^S.$$

Due to page limitation we omit the proofs of Lemma 11 and Lemma 12. In order to state the main theorem in full generality, we need to replace the fatness condition with a similar d-dimensional condition. In the d-dimensional case, we do not require that the limit in the capacity definition will exists but we need some notion stability in $\otimes_{i\in[d]}\Gamma_i$. This stability is presented in the following definition.

Definition 13. *For $\epsilon > 0$ and for $\Gamma \subseteq \mathcal{P}(\Sigma^{F_k^d})$, a d-axial product SCS, Γ is called* robust *if*

$$\overline{\mathsf{cap}}(\Gamma) = \mathsf{cap}(\Gamma).$$

The robustness of a d-axial product is the requirement that a small change in the set of restrictions does not affect greatly on the capacity.

We now state the main results of this section. This result concerns the relation between the d-dimensional independence entropy and capacity.

Theorem 14. *Let $\Gamma_i \subseteq \mathcal{P}(\Sigma^{k_i})$ be d one-dimensional closed and convex SCS, then $h_{\mathrm{ind}}(\otimes_{i=0}^{d-1}\Gamma_i) \leqslant \overline{\mathsf{cap}}(\otimes_{i=0}^{d-1}\Gamma_i)$. If $\otimes_{i=0}^{d-1}\Gamma_i$ is also robust, then $h_{\mathrm{ind}}(\otimes_{i=0}^{d-1}\Gamma_i) \leqslant \mathsf{cap}(\otimes_{i=0}^{d-1}\Gamma_i)$.*

Due to page limitation, we provide a sketch of the proof for the somewhat simpler case in which $d = 1$.

Sketch of proof: Fix n and consider $\mu \in \overline{\mathcal{P}}_n(\Gamma)$. For $m \in \mathbb{N}$, construct a measure μ^m over Σ^{nm} by concatenating μ to itself m times. Now we note that in μ^m, the non-overlapping n-tuples are distributed i.i.d according to μ. Denote this frequency by $\hat{\text{fr}}$ and use Cramer's theorem to obtain an exponential decay rate of the event $\mu^m\left(w \in \Sigma^{nm} \;:\; \hat{\text{fr}}^n_w \notin \mathbb{B}_\epsilon(\overline{\mathcal{P}}_n(\Gamma))\right)$. Denote the event $A \triangleq \left\{w \in \Sigma^{nm} \;:\; \text{fr}^k_w \in \mathbb{B}_\epsilon(\Gamma)\right\}$. Now use Lemma 11 and Lemma 12 to obtain that for $w \in \Sigma^{nm}$, $\text{fr}^k_w \notin \mathbb{B}_\epsilon(\Gamma) \Rightarrow \text{fr}^n_w \notin \mathbb{B}_\epsilon(\overline{\mathcal{P}}_n(\Gamma))$. This calculation holds for every cyclic shift of w hence we use the union bound to obtain an exponential decay rate of the event $\mu^m(\Sigma^{nm} \setminus A)$. By a straightforward calculation we now obtain that $h_{\text{ind}}(\Gamma) \leqslant \overline{\mathsf{cap}}(\Gamma)$. ■

Theorem 15. *Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a one-dimensional closed and convex SCS. Then the d-axial product of Γ with itself, namely $\Gamma^{\otimes d}$, satisfies*

$$h_{\text{ind}}(\Gamma) \leqslant h_{\text{ind}}(\Gamma^{\otimes d}) \leqslant \overline{h_{\text{ind}}}(\Gamma).$$

Due to page limitation we provide a sketch of the proof.

sketch of proof: We first show that $h_{\text{ind}}(\Gamma^{\otimes d}) \leqslant \overline{h_{\text{ind}}}(\Gamma)$. Take $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$, which is a measure over $\Sigma^{F^d_n}$. Define $\mu_i \triangleq \delta^{[n]\mathbf{e}_0}_{\mathbf{v}_i}(\mu)$ with $\mathbf{v}_i$ ranging over $\{0\} \times F^{d-1}_n$ as i ranges over $[n^{d-1}]$. Thus, each μ_i is a measure over Σ^n. Now construct $\hat{\mu} \in \mathcal{P}(\Sigma^{n^d})$ as follows. For a word $a = (a_0 \dots a_{n^d-1}) \in \Sigma^{n^d}$,

$$\hat{\mu}(a) \triangleq \prod_{i \in [n^{d-1}]} \mu_i(a_{ni} \dots a_{ni+n-1}),$$

i.e., the measure $\hat{\mu}$ is the product of the measures μ_i. Since the μ_i measures are independent, the concatenation measure, $\hat{\mu} \in \mathcal{P}(\Sigma^{n^d})$ is independent as well. We now calculate and obtain that $\delta^{[k]\mathbf{e}_0}(\mu) - \frac{k}{n} \leqslant \delta^{[k]}(\hat{\mu}) \leqslant \delta^{[k]\mathbf{e}_0}(\mu) + \frac{k}{n}$. This means that $\delta^{[k]}(\hat{\mu}) \in \mathbb{B}_{\frac{2k}{n}}(\Gamma)$. We obtained that for every $\epsilon > 0$, we can find $n_0 \in \mathbb{N}$ such that for every $n > n_0$, for every $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$, there is $\hat{\mu} \in \overline{\mathcal{P}}_{n^d}(\mathbb{B}_\epsilon(\Gamma))$. Since μ and $\hat{\mu}$ are independent we have

$$\begin{aligned} H(\mu) &= \sum_{\mathbf{v} \in F^d_n} H(\delta^{[\mathbf{1}]}_{\mathbf{v}}(\mu)) \\ &= \sum_{i \in [n^d]} H(\delta^1_i(\hat{\mu})) \\ &= H(\hat{\mu}). \end{aligned}$$

This implies that for every $\epsilon > 0$,

$$\limsup_{n\to\infty} \sup_{\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})} \frac{1}{n^d} H(\mu) \leqslant \limsup_{n\to\infty} \sup_{\mu \in \overline{\mathcal{P}}_{n^d}(\mathbb{B}_\epsilon(\Gamma))} \frac{1}{n^d} H(\mu).$$

We therefore obtain $h_{\text{ind}}(\Gamma^{\otimes d}) \leqslant h_{\text{ind}}(\mathbb{B}_\epsilon(\Gamma))$ for every ϵ which means that

$$h_{\text{ind}}(\Gamma^{\otimes d}) \leqslant \overline{h_{\text{ind}}}(\Gamma).$$

We now show the other direction. Take $\hat{\mu} \in \overline{\mathcal{P}}_n(\Gamma)$, and then also $\sigma(\hat{\mu}) \in \overline{\mathcal{P}}_n(\Gamma)$, where $\sigma(\hat{\mu})$ is the left cyclic shift of $\hat{\mu}$. Further recall that $\hat{\mu}$ is independent, and therefore can be written as $\hat{\mu} = \prod_{i=0}^{n-1} \delta^1_i(\hat{\mu})$. We now construct a measure $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$ using $\hat{\mu}$. For every $\mathbf{u} \in F^d_n$ set

$$\delta^{[\mathbf{1}]}_{\mathbf{u}}(\mu) \triangleq \delta^1_{\|\mathbf{u}\|_1}(\hat{\mu}),$$

where $\|\mathbf{u}\|_1 \triangleq \sum_{i=0}^{d-1} |u_i|$ is the ℓ_1 norm. Observe that μ is such that on every row in every direction we obtain $\sigma^t(\hat{\mu})$ for some t. Thus, we obtain that $\mu \in \overline{\mathcal{P}}_n(\Gamma^{\otimes d})$ and $H(\mu) = n^{d-1} H(\hat{\mu})$. Since we are taking the supremum over all measures we have $h_{\text{ind}}(\Gamma) \leqslant h_{\text{ind}}(\Gamma^{\otimes d})$. ■

We mention here that a careful look at the proof reveals something slightly stronger. Since we may take μ_i to be in any direction and not necessarily with the direction of the 0 axis, we obtain that for d one-dimensional closed and convex SCSs, the independence entropy of the d-axial product satisfies

$$h_{\text{ind}}(\otimes_{i=0}^{d-1} \Gamma_i) \leqslant \min_{i \in [d]} \overline{h_{\text{ind}}}(\Gamma_i).$$

IV. Discussion

We discuss some of the results obtained in this paper and compare them with known results. Consider $\Gamma_0 \subseteq \mathcal{P}(\Sigma^k)$, a one dimensional SCS. There are two natural generalizations of one-dimensional SCS to d-dimensional SCS which we consider interesting. In the first generalization, we may define

$$\Gamma \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{F^d_k}) \;:\; \forall i \in [d],\ \delta^{[k]\mathbf{e}_i}_{\mathbf{0}} \in \Gamma_0\right\},$$

namely, the d-axial product $\Gamma \triangleq \Gamma_0^{\otimes d}$. On the other hand, we may define the d-dimensional SCS in which the *average* of the k-length marginals is in Γ_0, namely,

$$\Gamma' \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{F^d_k}) \;:\; \frac{1}{d} \sum_{i=0}^{d-1} \delta^{[k]\mathbf{e}_i}_{\mathbf{0}} \in \Gamma_0\right\}.$$

Clearly,

$$\mathsf{cap}(\Gamma) \leqslant \mathsf{cap}(\Gamma'),$$

since $\Gamma \subseteq \Gamma'$.

We now focus on a simple example known as the $(0,k,p)$-RLL SCS over the binary alphabet $\Sigma = \{0,1\}$. The one-dimensional $(0,k,p)$-RLL SCS, $0 \leqslant p \leqslant 1$, is defined by

$$\Gamma_0 \triangleq \left\{\mu \in \mathcal{P}(\Sigma^{k+1}) \;:\; \mu(1^{k+1}) \leqslant p\right\},$$

where 1^{k+1} denotes the all-ones string of length $k+1$. This example is a generalization of the well known inverted $(0,k)$-RLL fully constrained systems. If we take $p = 0$ we obtain the inverted $(0,k)$-RLL. Let Γ and Γ' be the two d-dimensional SCS described above, using the $(0,k,p)$-RLL SCS of Γ_0.

In [6], the authors found lower and upper bounds on the capacity of Γ', which we recall here:

Theorem 16. *[6, Th. 16, Th. 20] Let Γ' denote the d-dimensional $(0,k,p)$-RLL SCS. Then,*

$$\begin{aligned} \mathsf{cap}(\Gamma') &\leqslant 1 - \frac{d\frac{\log_2 e}{2^{k+1}} + pd(k+1) - pd\log_2\frac{e}{p}}{3 - 2^{1-k} + 2^{-k}(d-1)(k+1)^2}, \\ \mathsf{cap}(\Gamma') &\geqslant 1 + d\left(\mathsf{cap}(\Gamma_0) - 1\right). \end{aligned}$$

Note that for high dimensions, the lower bound will be negative, and therefore, degenerate. Although the d-dimensional $(0,k,p)$-RLL SCS Γ', is different from the d-axial-product SCS Γ, the capacity lower bound obtained on the latter also lower bounds the capacity of the former.

Example 17. *Let us take $k = 2$, and $p = 0.05$, meaning that we restrict the frequency of the pattern 111 to be at most 0.05. Fix $d = 3$, and consider the two 3-dimensional SCS discussed above, $\Gamma \triangleq \Gamma_0^{\otimes 3}$, and Γ'.*

The lower bound on $\mathsf{cap}(\Gamma')$ from [6] uses $\mathsf{cap}(\Gamma_0)$. The latter can be calculated by solving an optimization problem using a computer. We obtain that $\mathsf{cap}(\Gamma_0) \approx 0.976$ which means that

$$\mathsf{cap}(\Gamma') \geqslant 1 + 3 \cdot (0.976 - 1) \approx 0.928.$$

It is clear that $\overline{\mathsf{cap}}(\Gamma') \geqslant \mathsf{cap}(\Gamma')$. Using Theorem 14 we obtain

$$\overline{\mathsf{cap}}(\Gamma') \geqslant \overline{\mathsf{cap}}(\Gamma) \geqslant \overline{h_{\mathrm{ind}}}(\Gamma_0^{\otimes 3}) \geqslant \overline{h_{\mathrm{ind}}}(\Gamma_0) \geqslant h_{\mathrm{ind}}(\Gamma_0).$$

Since finding the supremum involved in the definition of $h_{\mathrm{ind}}(\Gamma_0)$ is hard, we lower bound it by guessing a specific measure. We take each coordinate to be i.i.d. Bernoulli $\sqrt[3]{0.05}$, and we get

$$\overline{\mathsf{cap}}(\Gamma') \geqslant h_{\mathrm{ind}}(\Gamma_0) \geqslant H(\sqrt[3]{0.05}) \approx 0.949,$$

which is a better lower bound than that of [6]. Note that the upper bound gives $\mathsf{cap}(\Gamma') \leqslant 0.983$.

We further mention that the lower bound of [6] gets increasingly worse as the dimension grows. For example, when $d = 10$ we obtain by Theorem 16 that $\overline{\mathsf{cap}}(\Gamma') \geqslant 0.76$ whereas using the independence entropy, the bound stays the same, i.e., $\overline{\mathsf{cap}}(\Gamma') \geqslant 0.949$. □

REFERENCES

[1] R. J. Baxter, "Hard hexagons: exact solution," *J. Phys. A: Math. Gen.*, vol. 13, pp. L61–L70, 1980.

[2] N. Calkin and H. Wilf, "The number of independent sets in the grid graph," *SIAM J. Discrete Math.*, vol. 11, pp. 54–60, 1998.

[3] Y. M. Chee, J. Chrisnata, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Efficient encoding/decoding of capacity-achieving constant-composition ICI-free codes," in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT2016), Barcelona, Spain*, Jul. 2016, pp. 205–209.

[4] ——, "Rates of constant-composition codes that mitigate intercell interference," in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT2016), Barcelona, Spain*, Jul. 2016, pp. 200–204.

[5] O. Elishco, T. Meyerovitch, and M. Schwartz, "Encoding semiconstrained systems," in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT2016), Barcelona, Spain*, Jul. 2016, pp. 395–399.

[6] ——, "Semiconstrained systems," *IEEE Trans. Inform. Theory*, vol. 62, no. 4, pp. 811–824, Apr. 2016.

[7] H. Follmer and S. Orey, "Large deviations for the empirical field of a Gibbs measure," *Ann. Probab.*, vol. 16, no. 3, pp. 961–977, Jul. 1988.

[8] 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, no. 5, pp. 824–838, May 2004.

[9] H. Ito, A. Kato, Z. Nagy, and K. Zeger, "Zero capacity region of multidimensional run length constraints," *Elec. J. of Comb.*, vol. 6, 1999.

[10] R. Karabed, D. Neuhoff, and A. Khayrallah, "The capacity of costly noiseless channels," IBM Research Report, Tech. Rep. RJ 6040 (59639), Jan. 1988.

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

[12] S. Kayser and P. H. Siegel, "Constructions for constant-weight ICI-free codes," in *Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT2014), Honolulu, HI, USA*, Jul. 2014, pp. 1431–1435.

[13] A. S. Khayrallah and D. L. Neuhoff, "Coding for channels with cost constraints," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 854–867, 1996.

[14] O. F. Kurmaev, "Constant-weight and constant-charge binary run-length limited codes," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4497–4515, Jul. 2011.

[15] E. H. Lieb, "Residual entropy of square ice," *Physical Review*, vol. 162, no. 1, pp. 162–172, 1967.

[16] E. Louidor, B. Marcus, and R. Pavlov, "Independence entropy of $\mathbb{Z}^d$-shift spaces," *Acta Applicandae Mathematicae*, vol. 126, no. 1, pp. 297–317, 2013.

[17] E. Louidor and B. H. Marcus, "Improved lower bounds on capacities of symmetric 2D constraints using Rayleigh quotients," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1624–1639, Apr. 2010.

[18] B. H. Marcus and R. M. Roth, "Improved Gilbert-Varshamov bound for constrained systems," *IEEE Trans. Inform. Theory*, vol. 38, no. 4, pp. 1213–1221, 1992.

[19] T. Meyerovitch and R. Pavlov, "On independence and entropy for high-dimensional isotropic subshifts," *Proceedings of the London Mathematical Society*, p. pdu029, 2014.

[20] Z. Nagy and K. Zeger, "Capacity bounds for the three-dimensional $(0, 1)$ run length limited channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 1030–1033, May 2000.

[21] E. Ordentlich and R. M. Roth, "Independent sets in regular hypergraphs and multi-dimensional runlength-limited constraints," *SIAM J. Discrete Math.*, vol. 17, no. 4, pp. 615–623, 2004.

[22] T. L. Poo, P. Chaichanavong, and B. Marcus, "Tradeoff functions for constrained systems with unconstrained positions," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1425–1449, Apr. 2006.

[23] M. Qin, E. Yaakobi, and P. H. Siegel, "Constrained codes that mitigate inter-cell interference in read/write cycles for flash memories," *IEEE J. Select. Areas Commun.*, vol. 32, no. 5, pp. 836–846, May 2014.

[24] M. Schwartz and J. Bruck, "Constrained codes as networks of relations," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2179–2195, May 2008.

[25] M. Schwartz and A. Vardy, "New bounds on the capacity of multidimensional run-length constraints," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4373–4382, Jul. 2011.

[26] A. Shafarenko, A. Skidin, and S. K. Turitsyn, "Weakly-constrained codes for suppression of patterning effects in digital communications," *IEEE Trans. Communications*, vol. 58, no. 10, pp. 2845–2854, Oct. 2010.

[27] A. Shafarenko, K. S. Turitsyn, and S. K. Turitsyn, "Information-theory analysis of skewed coding for suppression of pattern-dependent errors in digital communications," *IEEE Trans. Communications*, vol. 55, no. 2, pp. 237–241, Feb. 2007.

[28] A. Sharov and R. M. Roth, "Two-dimensional constrained coding based on tiling," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1800–1807, Apr. 2010.

[29] P. H. Siegel and J. K. Wolf, "Bit-stuffing bounds on the capacity of 2-dimensional constrained arrays," in *Proceedings of the 1998 IEEE International Symposium on Information Theory (ISIT1998), Cambridge, MA, USA*, Aug. 1998, p. 323.

[30] I. Tal and R. M. Roth, "Bounds on the rate of 2-D bit-stuffing encoders," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2561–2567, Jun. 2010.