On Encoding Semiconstrained Systems

Ohad Elishco, *Student Member, IEEE*, Tom Meyerovitch, and Moshe Schwartz, *Senior Member, IEEE*

***Abstract*— Semiconstrained systems (SCSs) were recently suggested as a generalization of constrained systems, commonly used in communication and data-storage applications that require certain offending subsequences be avoided. In an attempt to apply the techniques from constrained systems, we study the sequences of constrained systems that are contained in, or contain, a given SCS, while approaching its capacity. In the former case, we describe two such sequences resulting in constant-to-constant bit-rate block encoders and finite-state encoders. Perhaps surprisingly, we show in the latter case, under commonly made assumptions, that the only constrained system that contains a given SCS is the entire space. A refinement to this result is also provided, in which semiconstraints and zero constraints are mixed together.**

***Index Terms*— Constrained coding, channel capacity, encoding.**

I. Introduction

MANY communication and data-storage systems employ constrained coding. In such a scheme, information is encoded in sequences that avoid the occurrence of certain subsequences. Perhaps the most common example is that of (d, k)-RLL which is comprised of binary sequences that avoid subsequences of $k+1$ 0's, as well as two 1's that are separated by less than d 0's. For various other examples the reader is referred to [7] and the many references therein.

The reason for avoiding such subsequences is mainly due to the fact that their appearance contributes to noise in the system. However, by altogether forbidding their occurrence, the possible rate at which information may be transmitted is severely reduced. By relaxing the constraints and allowing some appearances of the offending subsequences, the rate penalty may be reduced. So rather than imposing combinatorial constraints on substrings, we impose statistical constraints on them. Such an approach was studied, for example, in the case of channels with cost constraints [8], [10].

A general approach was suggested in [5], in which a *semiconstrained system (SCS)* was defined by a list of offending subsequences, and an upper bound (called a *semiconstraint*) on the frequency of each subsequence appearing. Note that constrained systems (which we call *fully constrained* for emphasis) are a special case of semiconstrained systems, in which only semiconstraints of frequency 0 are used.

A careful choice of semiconstraints also allows the study of systems that, up to now, were studied in an ad-hoc manner only. As examples we mention DC-free RLL coding [12], constant-weight ICI coding for flash memories [2], [3], [9], [14], and coding to mitigate the appearance of ghost pulses in optical communication [15], [16].

One of the most important questions, given a SCS, is how to encode any unconstrained input sequence into a sequence that satisfies all the given semiconstraints. The various encoding schemes suggested in [2], [9], [12], and [14]–[16] are ad-hoc and do not apply to general SCSs. The encoding scheme for channels with cost constraints given in [10] (which overlap somewhat with SCSs) is indeed general, however it is not capacity achieving. Later, within the scope of channels with cost constraints, and motivated by partial-response channels, [11], [19] briefly report on capacity-achieving schemes, however, not in the full generality we consider in this paper.

Under the assumption that the input stream consists of i.i.d. uniformly-random bits, a general capacity-achieving encoding scheme for SCSs was described in [5]. The scheme involved a maxentropic Markov chain over a modified De-Bruijn graph. Input symbols were converted via an arithmetic decoder to a biased stream of symbols which were used to generate a path in the graph, which in turn generated symbols to be transmitted. A reverse operation was employed at the receiving side. Additionally to the assumption on the distribution of the input, to enforce a constant-to-constant bit rate, the encoder has a probability of failure (albeit, asymptotically vanishing). Thus, not all input streams may be converted to semiconstrained sequences.

Compared with SCSs, for "conventional" fully constrained systems there is a general method for constructing encoders working arbitrarily close to capacity: the celebrated state-splitting algorithm. However, as we explain in the following sections, this method fails even on very simple SCSs, due to the fact that in most cases they do not form regular languages.

In this work we consider the problem of encoding an arbitrary input string into a sequence that satisfies all the given semiconstraints. We do not make statistical or combinatorial assumptions on the input, only that it is sufficiently long. Specifically, we show the following: For every given SCS that satisfies certain mild assumptions and every $\epsilon > 0$ we present a fully constrained system that is "eventually-contained" in the given SCS, with a capacity penalty of at most ϵ. This allows us to construct either block encoders or finite-state encoders, trading encoder anticipation for number of states. In the

Manuscript received October 7, 2016; revised August 13, 2017; accepted October 5, 2017. Date of publication November 8, 2017; date of current version March 15, 2018. This work was supported in part by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant 333598 and in part by the Israel Science Foundation under Grant 626/14. This paper was presented in part at the 2016 IEEE International Symposium on Information Theory.

O. Elishco and M. Schwartz are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (e-mail: ohadeli@post.bgu.ac.il; schwartz@ee.bgu.ac.il).

T. Meyerovitch is with the Department of Mathematics, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (e-mail: mtom@math.bgu.ac.il).

Communicated by A. Ramamoorthy, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2017.2771743

other direction, we show that no fully constrained system can contain a given SCS (under certain mild assumptions).

The paper is organized as follows. In Section II we present the definition and notation used throughout the paper. In Section III we study sequences of constrained systems that are contained in a given SCS and approach its capacity from below. In Section IV we do the reverse, and study constrained systems containing a given SCS. In the first sections of this paper we make some assumptions on the SCS under discussion, in particular that they are *fat* (see Definition 7). These assumptions in particular exclude (classical) fully constrained systems. In Section V we study more general SCSs, that in particular also allow fully constrained systems. We present conclusions and further research in Section VI.

II. Preliminaries

A. Semiconstrained Systems

Let Σ be a finite alphabet and let Σ^* denote the set of all the finite sequences over Σ. The elements of Σ^* are called *words* (or *strings*). The *length* of a word $\omega \in \Sigma^*$ is denoted by $|\omega|$. Given two words, $\omega, \omega' \in \Sigma^*$, their concatenation is denoted by $\omega\omega'$. Repeated concatenation is denoted using a superscript, i.e., for any natural $m \in \mathbb{N}$, ω^m denotes $\omega^m = \omega\omega\ldots\omega$, where m copies of ω are concatenated. By convention, $\omega^0 = \varepsilon$, where ε the unique empty word of length 0. By extension, if $S \subseteq \Sigma^*$ is a set of words, then S^m denotes the set

$$S^m = \{\omega_1\omega_2\ldots\omega_m : \forall i, \omega_i \in S\},$$

with $S^0 = \{\varepsilon\}$, $S^* = \bigcup_{i\geqslant 0} S^i$, and $S^+ = \bigcup_{i\geqslant 1} S^i$.

The set of length-k subwords of ω is defined by

$$\mathrm{sub}_k(\omega) = \left\{\beta \in \Sigma^k \ : \ \omega = \alpha\beta\gamma \ \text{ for some } \alpha, \gamma \in \Sigma^*\right\}.$$

For $\omega \in \Sigma^*$ and $k \leqslant |\omega|$, $\mathrm{fr}^k_\omega \in \mathcal{P}(\Sigma^k)$ is defined as the empirical distribution of length-k subwords in ω, where we denote by $\mathcal{P}(\Sigma^k)$ the set of all probability measures on Σ^k. We can naturally identify

$$\mathcal{P}(\Sigma^k) = \left\{\eta \in [0,1]^{\Sigma^k} : \sum_{\phi\in\Sigma^k} \eta(\phi) = 1\right\}.$$

It follows that for all $\beta \in \Sigma^k$,

$$\mathrm{fr}^k_\omega(\beta) = \frac{1}{|\omega| - k + 1}\left|\left\{(\alpha,\gamma) : \alpha,\gamma \in \Sigma^*, \alpha\beta\gamma = \omega\right\}\right|.$$

Example 1: Let $\Sigma = \{0,1\}$ and let $\omega = 0110100100 \in \Sigma^{10}$. Then $\mathrm{fr}^2_\omega(01) = \frac{3}{9}$. □

Definition 2: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a set of probability measures. We say Γ is a semiconstrained system (SCS)*, and we define the set of* admissible words for Γ *by*

$$\mathcal{B}(\Gamma) = \left\{\omega \in \Sigma^* : \mathrm{fr}^k_\omega \in \Gamma\right\}.$$

For convenience we also define the set of admissible words of length exactly n as

$$\mathcal{B}_n(\Gamma) = \mathcal{B}(\Gamma) \cap \Sigma^n.$$

Example 3: We consider a family of SCSs known as the $(0,1,p)$-RLL SCS [5]. Let $\Sigma = \{0,1\}$, and let $p \in [0,1]$. Define

$$\Gamma = \left\{\eta \in \mathcal{P}(\Sigma^2) \ : \ \eta(11) \leqslant p\right\}.$$

The set of admissible words for Γ is the set of all binary words whose empirical frequency of the pattern 11 is at most p. Taking $p = 0$ we obtain the well known $(1,\infty)$-RLL fully constrained system. □

An important figure of merit we associate with any set of words $S \subseteq \Sigma^*$ is its capacity.

Definition 4: Let Σ be a finite alphabet and $S \subseteq \Sigma^*$. The capacity *of S, denoted $\mathsf{cap}(S)$, is defined as*

$$\mathsf{cap}(S) = \limsup_{n\to\infty} \frac{1}{n}\log_2\left|S \cap \Sigma^n\right|.$$

Thus, in the case of a SCS Γ, the capacity $\mathsf{cap}(\mathcal{B}(\Gamma))$ intuitively measures the exponential growth rate of the number of words that satisfy the constraints given by Γ as a function of the word length.

A relaxation of semiconstrained systems was also suggested in [5].

Definition 5: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a set of probability measures. For $\epsilon > 0$ we denote by Γ^ϵ the set

$$\Gamma^\epsilon = \left\{\eta \in \mathcal{P}(\Sigma^k) \ : \ \inf_{\mu\in\Gamma} \|\eta - \mu\|_\infty \leqslant \epsilon\right\}$$

where $\|\cdot\|_\infty$ is the ℓ_∞-norm. The set of weakly-admissible words for Γ *is defined by*

$$\overline{\mathcal{B}}(\Gamma) = \left\{\omega \in \Sigma^* : \mathrm{fr}^k_\omega \in \Gamma^{\xi(|\omega|)}\right\},$$

where $\xi : \mathbb{N} \to \mathbb{R}^+$ is a function satisfying both $\xi(n) = o(1)$ and $\xi(n) = \Omega(1/n)$. Also $\overline{\mathcal{B}}_n(\Gamma) = \overline{\mathcal{B}}(\Gamma) \cap \Sigma^n$.

Note that the term $\zeta(|\omega|)$ adds some tolerance. In order for that tolerance to be negligible in terms of capacity, we require $\zeta(n) = o(1)$. In order for it to be effective enough to solve continuity issues, we also require $\zeta(n) = \Omega(\frac{1}{n})$. Intuitively, the latter requirement allows us to alter a constant number of letters in a string of any length. Any tolerance of order $o(\frac{1}{n})$ is eventually degenerate, and equivalent to setting $\zeta(n) = 0$.

We note that $\overline{\mathcal{B}}(\Gamma)$ was called a *weak semiconstrained system (WSCS)* in [5] and was defined in a slightly different manner, though we shall prefer to use the term weakly-admissible words for Γ. It was also shown there that though it is possible to constrain words of different lengths, it suffices to consider only words in Σ^k, i.e., all the offending patterns are of the same length k. Here, since $\Gamma \subseteq \mathcal{P}(\Sigma^k)$, the assumption that all the offending patterns are of length k is implied.

A particular set of probability measures of interest to us is the set of *shift-invariant probability measures*. We say $\eta \in \mathcal{P}(\Sigma^k)$ is *shift-invariant* if for all $\phi \in \Sigma^{k-1}$,

$$\sum_{a\in\Sigma} \eta(a\phi) = \sum_{a\in\Sigma} \eta(\phi a).$$

We denote the set of shift-invariant probability measures by $\mathcal{P}_{\mathrm{si}}(\Sigma^k)$, which is a closed subset of $\mathcal{P}(\Sigma^k)$ in the weak-* topology (since Σ^k is a finite set with the discrete topology, the topology on $\mathcal{P}(\Sigma^k)$ is given by the total variation norm). These are precisely the probability measures that arise as marginals of shift-invariant measures on $\Sigma^{\mathbb{N}}$ or $\Sigma^{\mathbb{Z}}$. For a discussion see [1]. In particular, we have the following lemma.

Lemma 6: Fix a finite alphabet Σ, and $k \geqslant 2$. If $\Gamma \subseteq \mathcal{P}(\Sigma^k) \setminus \mathcal{P}_{\text{si}}(\Sigma^k)$ is closed then $\mathsf{cap}(\mathcal{B}(\Gamma)) = -\infty$, i.e., $\mathcal{B}(\Gamma)$ is a finite set.

Proof: For any $\omega \in \Sigma^*$, $|\omega| \geqslant k$, and any $\phi \in \Sigma^{k-1}$, by simple counting

$$\left| \sum_{a\in\Sigma} \mathrm{fr}^k_\omega(a\phi) - \sum_{a\in\Sigma} \mathrm{fr}^k_\omega(\phi a) \right| \leqslant \frac{1}{|\omega| - k + 1}.$$

Thus, fr^k_ω gets arbitrarily close to a shift-invariant probability measure as $|\omega| \to \infty$. Since $\mathcal{P}_{\text{si}}(\Sigma^k)$ and Γ are closed, there is a positive distance between the sets. Therefore, there exists $n \in \mathbb{N}$ such that for all $\omega \in \Sigma^*$, $|\omega| \geqslant n$, we have $\omega \notin \mathcal{B}(\Gamma)$, i.e., $\mathcal{B}(\Gamma)$ is finite. ■

Lemma 6 motivates us to study probability measures that are shift invariant. Another crucial property of a set of probability measures is given in the following definition.

Definition 7: For a set $\Gamma' \subseteq \mathcal{P}_{\text{si}}(\Sigma^k)$ we denote by $\mathrm{int}(\Gamma')$ *the interior of Γ', and by* $\mathrm{cl}(\Gamma')$ *the closure of Γ', both relatively to $\mathcal{P}_{\text{si}}(\Sigma^k)$. We say $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ is* fat *if*

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

The fat condition allows us to focus on "well-behaved" sets, namely, sets with a non-empty interior. the following result from [5] demonstrates the importance of a fat Γ.

Theorem 8: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be closed and convex. If Γ is fat then

$$\mathsf{cap}(\mathcal{B}(\Gamma)) = \mathsf{cap}(\overline{\mathcal{B}}(\Gamma)) = \log_2 |\Sigma| - \inf_{\eta\in\Gamma\cap\mathcal{P}_{\text{si}}(\Sigma^k)} D(\eta \| \eta'),$$

where $D(\cdot\|\cdot)$ is the relative entropy function, and $\eta'(\phi a) = \sum_{a'\in\Sigma} \eta(\phi a')/|\Sigma|$, for all $\phi \in \Sigma^{k-1}$ and $a \in \Sigma$. Additionally, $\mathsf{cap}(\mathcal{B}(\Gamma))$ *and* $\mathsf{cap}(\overline{\mathcal{B}}(\Gamma))$ *are continuous and convex in Γ, and the limits in their definitions exist.*

B. Fully Constrained Systems

As noted in the introduction, "conventional" constrained systems are a special case of semiconstrained systems. A constrained system can be viewed as a SCS Γ, where Γ is of the form

$$\Gamma = \left\{ \eta \in \mathcal{P}(\Sigma^k) : \forall \phi \in \Sigma^k, \eta(\phi) \leqslant c_\phi \right\},$$

where $c_\phi \in \{0,1\}$ for all $\phi \in \Sigma^k$. In other words, every substring of length k is either completely forbidden, or unconstrained. We will refer to those as *fully constrained systems* and denote a set Γ of this form as $\Gamma_{\{0,1\}}$.

Let $G = (V, E)$ be a finite directed graph, where we allow parallel edges. A labeling function $\mathcal{L} : E \to \Sigma^q$ assigns a length-q label over the alphabet to each edge. By simple extension, the label of a directed (non-empty) path in the graph $\gamma = e_1 \to e_2 \to \cdots \to e_n$ is defined as $\mathcal{L}(\gamma) = \mathcal{L}(e_1)\mathcal{L}(e_2)\dots\mathcal{L}(e_n)$. Finally, we define the language represented by the graph G, denoted $\mathcal{L}(G)$, to be the labels of all finite directed paths in G.

Constrained systems have been widely studied [7], [13]. In particular, it is well known that in case Γ is of the form $\Gamma_{\{0,1\}}$, $\mathcal{B}(\Gamma) = \mathcal{L}(G)$ for some finite directed labeled graph G in the manner described above. An immediate consequence is the fact that $\mathcal{B}(\Gamma)$ is a regular language in the Chomsky hierarchy of formal languages [17]. We do note, however, that not all regular languages (which correspond to languages of sofic subshifts) are constrained systems (which are defined by a finite number of forbidden words, and correspond to subshifts of finite type).

A wide variety of tools exist for manipulating constrained systems, including the state-splitting algorithm (see [13, Ch. 5]). In essence, under mild assumptions, given a constrained system $\mathcal{B}(\Gamma) = \mathcal{L}(G)$, and two positive integers p and q that satisfy $p/q < \mathsf{cap}(\mathcal{B}(\Gamma))$, we can find another constrained system $\mathcal{B}'(\Gamma) = \mathcal{L}(G')$, an *encoder*, with the following properties:

- $\mathcal{L}(G') \subseteq \mathcal{L}(G)$.
- $\mathsf{cap}(\mathcal{B}'(\Gamma)) = p/q$, also called the *rate* of the encoder.
- G' is a $p : q$ encoder for $\mathcal{L}(G)$ with finite anticipation $a \in \mathbb{N} \cup \{0\}$, i.e., the out-degree of each vertex is 2^p, the edge labels in G' are from Σ^q, and paths of length $a + 1$ that start from the same vertex and generate the same word agree on the first edge.

Unfortunately, even for very simple semiconstraints, $\mathcal{B}(\Gamma)$ is not a regular language in general. As an example, for $\Sigma = \{0, 1\}$, and Γ such that for all $\mu \in \Gamma$, $\mu(1) \leqslant p$, it is easily seen that for any rational $0 < p < 1$, the semiconstrained system $\mathcal{B}(\Gamma)$ is a non-regular context-free language, whereas for any irrational p the system is not even context free [17, Sec. 4.9, Exercise 25]. Thus, the wonderful machinery of the state-splitting algorithm cannot be applied directly for general SCSs.

Another important property of languages associated with fully constrained systems is that these languages are *factorial*. This means that a subword of an admissible word is also an admissible word. Factoriality implies for instance that if Γ is of the form $\Gamma_{\{0,1\}}$, the sequence $\frac{1}{n} \log \left|\mathcal{B}(\Gamma_{\{0,1\}})\right|$ is subadditive, so the lim sup in the definition of the capacity is actually a limit by Fekete's Lemma. The factoriality property is not shared by SCSs in general.

III. Approaching Capacity From Below

In this section we study the problem of finding a sequence of fully constrained systems that are contained in a given semiconstrained (or weakly semiconstrained) system, with the additional requirement that the capacity of the former approaches that of latter in the limit. We present two such sequences which induce (perhaps after state splitting) two possible encoders for the SCS or WSCS.

Before continuing on, we pause to consider what properties we require of an encoder. An encoder is nothing more than a function $\phi : \Sigma^{\mathbb{N}} \to X$ for translating an unconstrained sequence of input symbols $\Sigma^{\mathbb{N}}$, into another sequence obeying a given set of constraints, $X \subseteq \Sigma^{\mathbb{N}}$. A general encoder for SCSs was already described in [5]. However, that encoder had a probability of failure, i.e., it would not work on some input sequences. We are therefore interested in finding an encoder that always succeeds.

Thus, in what follows, we focus on studying fully constrained systems contained in a given SCS. One of our goals is to determine the following function.

Definition 9: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a SCS. We define

$$\mathsf{cap}^{\subseteq}(\Gamma) = \sup_{\mathcal{L}(G)\subseteq\mathcal{B}(\Gamma)} \mathsf{cap}(\mathcal{L}(G)).$$

We remark that this definition is superficially reminiscent of inner measures from measure theory. The definition is also provisional, since as we shall later show, it coincides with the usual notion of capacity.

It will be easier for us to describe fully constrained systems that are only *eventually* contained in the desired SCS. Formally, given two infinite subsets, $S_1, S_2 \in \Sigma^*$, we say S_1 is eventually contained in S_2, denoted $S_1 \subseteq^e S_2$, if $|S_1 \setminus S_2| < \infty$. A fully constrained system that is eventually contained in a given SCS may easily be transformed into another fully constrained system that is contained (in the usual sense) in the given SCS by removing the words that are inadmissible in the SCS.

A. Block Encoders for SCSs

The first sequence of fully constrained systems we construct are each represented by a graph with a single state. Such graphs are called block encoders.

Let Γ be a fat SCS. The fat condition on Γ guarantees that it can be slightly shrunk while remaining not empty. More formally, for any $\epsilon > 0$ we define the set Γ_ϵ by

$$\Gamma_\epsilon = \left\{ \eta \in \mathcal{P}(\Sigma^k) \ : \ \inf_{\mu\in\mathcal{P}(\Sigma^k)\setminus\Gamma} \|\eta - \mu\|_\infty > \epsilon \right\} \quad (1)$$

where $\|\cdot\|_\infty$ is the ℓ_∞-norm.

First note that for every $\epsilon > 0$, $\Gamma_\epsilon \subseteq \Gamma$. If Γ is fat then there exists $\epsilon > 0$ such that $\Gamma_\epsilon \neq \emptyset$ and Γ_ϵ is also fat. We say such an ϵ is Γ-*feasible*.

Note also that in the definition of Γ_ϵ we consider $\mu \in \mathcal{P}(\Sigma^k)$ as a vector of numbers and use the ℓ_∞-norm instead of the usual total-variation norm. The particular choice of norm is a side issue and does not significantly change the essential results.

Construction A: Let Γ be a SCS. For every $m \in \mathbb{N}$ we construct $R_m(\Gamma) \subseteq \Sigma^*$ by defining

$$R_m(\Gamma) = \mathcal{B}_m(\Gamma)^*.$$

□

By definition, $R_m(\Gamma)$ from Construction A is a regular language. It may be represented as the language of the following graph G: the graph contains a single vertex, all the edges are self loops and are labeled by the words of $\mathcal{B}_m(\Gamma)$, i.e., the length-m words in $\mathcal{B}(\Gamma)$.

The next theorem ties $R_m(\Gamma_\epsilon)$ with $\mathcal{B}(\Gamma)$. In order the prove this theorem however, we need a simple lemma first which shows that if Γ is convex then for any $\epsilon > 0$, Γ_ϵ is convex as well.

Lemma 10: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a convex set, then for any $\epsilon > 0$, Γ_ϵ is also convex.

Proof: Consider Γ as a subset of $V = [0,1]^{|\Sigma|^k}$. For $x \in V$, denote by $\{x\}^\epsilon$ the ϵ neighborhood of x, i.e., $\{x\}^\epsilon = \{z \in V \ : \ \|z - x\|_\infty \leqslant \epsilon\}$. The lemma follows from the observation that if $x, y \in \Gamma$ such that $\{x\}^\epsilon, \{y\}^\epsilon \subseteq \Gamma$, then for every $t \in [0,1]$, $\{tx + (1-t)y\}^\epsilon \subseteq \Gamma$. Indeed, let $x, y \in \Gamma$ such that $\{x\}^\epsilon, \{y\}^\epsilon \subseteq \Gamma$. Let $t \in [0,1]$ and denote by z the point $z = tx + (1-t)y$. Let $u \in \{z\}^\epsilon$, and write $u = z + v$ where $\|v\|_\infty \leqslant \epsilon$. We have that $x + v, y + v \in \Gamma$ and $t(x+v) + (1-t)(y+v) = z + v \in \Gamma$ since Γ is convex. The lemma follows by noticing that if $x \in \Gamma$ is such that $\{x\}^\epsilon \subseteq \Gamma$ then $x \in \Gamma_\epsilon$. ■

Theorem 11: Let Γ be a convex fat SCS. Then for any Γ-feasible $\epsilon > 0$, there exists $M_\epsilon \in \mathbb{R}$ such that for all $m > M_\epsilon$

$$R_m(\Gamma_\epsilon) \subseteq \mathcal{B}(\Gamma).$$

Proof: By Lemma 10, if Γ is convex then so is Γ_ϵ. Consider $\omega \in R_m(\Gamma_\epsilon)$ and write $\omega = \omega_1\omega_2\dots\omega_\ell$ with $\omega_i \in \mathcal{B}_m(\Gamma_\epsilon)$. For any $\phi \in \Sigma^k$, we bound the number of occurrences of ϕ in ω.

For every $1 \leqslant i \leqslant \ell$, we denote

$$\mu_i = \mathrm{fr}^k_{\omega_i} \in \Gamma_\epsilon.$$

Every $\phi \in \Sigma^k$ appears in ω_i exactly $\mu_i(\phi)(m-k+1)$ times. Additionally, in each concatenation point between ω_i and ω_{i+1}, the word ϕ can appear at most another $k-1$ times.

Since Γ_ϵ is convex, we have

$$\mu = \frac{1}{\ell}\sum_{i=1}^{\ell} \mu_i \in \Gamma_\epsilon.$$

It now follows that

$$\begin{aligned} \mathrm{fr}^k_\omega(\phi) &\leqslant \frac{\sum_{i=1}^{\ell} \mu_i(\phi)(m-k+1) + (\ell-1)(k-1)}{m\ell - k + 1} \\ &= \frac{\mu(\phi)\ell(m-k+1) + (\ell-1)(k-1)}{m\ell - k + 1} \\ &= \mu(\phi) + \frac{(\ell-1)(k-1)(1-\mu(\phi))}{m\ell - k + 1} \\ &\leqslant \mu(\phi) + \frac{(\ell-1)(k-1)}{m\ell - k + 1}. \end{aligned} \quad (2)$$

Additionally,

$$\begin{aligned} \mathrm{fr}^k_\omega(\phi) &\geqslant \frac{\sum_{i=1}^{\ell} \mu_i(\phi)(m-k+1)}{m\ell - k + 1} \\ &= \frac{\mu(\phi)\ell(m-k+1)}{m\ell - k + 1} \\ &= \mu(\phi) - \frac{(\ell-1)(k-1)}{m\ell - k + 1}. \end{aligned} \quad (3)$$

Following the right-hand side of (2) and of (3) we can continue the analysis assuming all ω_i share the same measure $\mu \in \Gamma_\epsilon$. Now define

$$L(m,\ell) = \frac{(\ell-1)(k-1)}{m\ell - k + 1},$$

and thus,

$$\left|\mathrm{fr}^k_\omega(\phi) - \mu(\phi)\right| \leqslant L(m,\ell).$$

We note that for every $m > k-1$, $\ell \geqslant 2$, the function $L(m,\ell)$ is monotone increasing in ℓ. It follows that

$$L(m,\ell) < \lim_{\ell\to\infty} L(m,\ell) = \frac{k-1}{m}.$$

Hence, we can take

$$M_\epsilon = \frac{k-1}{\epsilon}, \tag{4}$$

and obtain that for every $m > M_\epsilon$,

$$L(m, \ell) < \epsilon.$$

The calculation holds for every $\phi \in \Sigma^k$. Thus, we showed that for every $m > M_\epsilon$, every $\phi \in \Sigma^k$, and every $\omega \in R_m(\Gamma_\epsilon)$, we have $\mathrm{fr}_\omega^k \in \Gamma$, and therefore $R_m(\Gamma_\epsilon) \subseteq \mathcal{B}(\Gamma)$. ∎

We observe that $M_\epsilon = \Omega(\frac{1}{\epsilon})$. The following theorem shows that the sequence of systems $R_m(\Gamma_\epsilon)$ has a capacity that approaches $\mathsf{cap}(\mathcal{B}(\Gamma_\epsilon))$ as m grows.

Theorem 12: Let Γ be a closed convex fat SCS. Then for every Γ-feasible $\epsilon > 0$ the following limit exists and

$$\lim_{m\to\infty} \mathsf{cap}(R_m(\Gamma_\epsilon)) = \mathsf{cap}(\mathcal{B}(\Gamma_\epsilon)).$$

Proof: We observe that

$$\left|R_m(\Gamma_\epsilon) \cap \Sigma^n\right| = \begin{cases} |\mathcal{B}_m(\Gamma_\epsilon)|^{\frac{n}{m}} & \text{if } m|n, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned} \mathsf{cap}(R_m(\Gamma_\epsilon)) &= \limsup_{n\to\infty} \frac{1}{n} \log_2 \left|R_m(\Gamma_\epsilon) \cap \Sigma^n\right| \\ &= \frac{1}{m} \log_2 |\mathcal{B}_m(\Gamma_\epsilon)|. \end{aligned}$$

Hence,

$$\begin{aligned} \limsup_{m\to\infty} \mathsf{cap}(R_m(\Gamma_\epsilon)) &= \limsup_{m\to\infty} \frac{1}{m} \log_2 |\mathcal{B}_m(\Gamma_\epsilon)| \\ &= \mathsf{cap}(\mathcal{B}(\Gamma_\epsilon)), \end{aligned}$$

by the definition of capacity. However, since ϵ is Γ-feasible, we have a fat Γ_ϵ. By Theorem 8, the limit in the definition of the capacity for the SCS exists, which completes the proof. ∎

We note that if $\epsilon_1 \leqslant \epsilon_2$ and ϵ_2 is Γ-feasible, then ϵ_1 is also Γ-feasible.

Corollary 13: For any SCS with a closed convex fat Γ there exist block encoders with rate arbitrarily close to $\mathsf{cap}(\mathcal{B}(\Gamma))$.

Proof: By Theorem 8 the limit in the capacity definition exists and the capacity, which is given by the relative entropy function, is continuous with respect to the restrictions. Thus,

$$\lim_{\epsilon\to 0} \mathsf{cap}(\mathcal{B}(\Gamma_\epsilon)) = \mathsf{cap}(\mathcal{B}(\Gamma)),$$

where ϵ is Γ-feasible. It follows that Theorem 11 and Theorem 12 show that it is possible to build a block encoder to a given SCS with rate arbitrarily close to $\mathsf{cap}(\mathcal{B}(\Gamma))$. ∎

While the block encoders we constructed are quite simple, and have rate p/q arbitrarily close to $\mathsf{cap}(\mathcal{B}(\Gamma))$, we do however point a major drawback. The edges are labeled by words from Σ^m. Thus, the encoder is not $p : q$ but $mp : mq$. For a fair comparison with the next construction, if we transform this to an encoder with labels from Σ (e.g., via a standard Moore co-form and tree argument), the number of states becomes exponential in m, and the anticipation becomes $\Omega(m)$, which is undesirable.

B. Finite-State Encoders

Unlike Construction A, in which a sequence was a concatenation of independent blocks, the construction we now present has a sliding-window restriction.

Construction B: Let Γ be a SCS. For every $m \in \mathbb{N}$ we construct $N_m(\Gamma) \subseteq \Sigma^*$ by defining

$$N_m(\Gamma) = \left\{\omega \in \Sigma^* : \mathrm{sub}_m(\omega) \subseteq \mathcal{B}(\Gamma)\right\}.$$

□

We observe that $N_m(\Gamma)$ from Construction B is a fully constrained system. Indeed, it is defined by a finite set of forbidden words, $\Sigma^m \setminus \mathcal{B}_m(\Gamma)$.

For the purpose of building an encoder, we construct a labeled graph G that represents $N_m(\Gamma)$. The vertex set is defined as $V = \bigcup_{i=0}^{m-1} \Sigma^i$. The edges, with labels from Σ, are given by

$$a_0 a_1 \dots a_i \xrightarrow{a_{i+1}} a_0 a_1 \dots a_i a_{i+1},$$

for all $0 \leqslant i \leqslant m-2$ and $a_j \in \Sigma$ for all j, as well as

$$a_0 \; a_1 \dots a_{m-2} \xrightarrow{a_{m-1}} a_1 \dots a_{m-2} a_{m-1},$$

for all $a_0 a_1 \dots a_{m-2} a_{m-1} \in \mathcal{B}(\Gamma)$ and $a_j \in \Sigma$ for all j.

It is easily observed that every path of length $m-1$ labeled by $\omega \in \Sigma^{m-1}$ ends in the vertex labeled by ω. From then on, by simple induction, assuming $\omega'\omega$ is a label of a path with $\omega \in \Sigma^{m-1}$, then the path ends in the vertex ω and a letter $a \in \Sigma$ may be generated following that path if and only if $\omega a \in \mathcal{B}(\Gamma)$.

Theorem 14: Let Γ be a convex fat SCS. Then for any Γ-feasible $\epsilon > 0$, and for all $m \geqslant k$,

$$N_m(\Gamma_\epsilon) \subseteq^e \mathcal{B}(\Gamma).$$

Proof: Consider $\omega \in N_m(\Gamma_\epsilon)$, $\omega = a_1 \; a_2 \dots a_n$, $a_i \in \Sigma$, and assume $|\omega| = n \geqslant 3m-2$. We define the ith length-m window sliding over ω as

$$\omega_i = a_i a_{i+1} \dots a_{i+m-1},$$

for all $1 \leqslant i \leqslant n-m+1$. We conveniently denote

$$\mu_i = \mathrm{fr}_{\omega_i}^k \in \Gamma_\epsilon.$$

We also define

$$\mu = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \mu_i.$$

Since Γ is convex, so is Γ_ϵ, and therefore $\mu \in \Gamma_\epsilon$.

For any $\phi \in \Sigma^k$, the number of occurrences of ϕ in ω_i is exactly $(m-k+1)\mu_i(\phi)$. By taking the sum $(m-k+1) \sum_{i=1}^{n-m+1} \mu_i(\phi)$ we are overcounting the number of times ϕ occurs in ω. However, we note that any occurrence of ϕ that is fully contained within $a_m a_{m+1} \dots a_{n-m+1}$ (i.e., within the windows $\omega_m, \omega_{m+1}, \dots, \omega_{n-2m+2}$), is overcounted by a factor of $m-k+1$ since it appears within exactly $m-k+1$ consecutive length-m windows ω_i. It follows that

$$\begin{aligned}\mathrm{fr}_{\omega}^{k}(\phi) &\leqslant \frac{1}{n-k+1}\sum_{i=m}^{n-2m+2}\mu_i(\phi)\\&\quad+\frac{m-k+1}{n-k+1}\left(\sum_{i=1}^{m-1}\mu_i(\phi)+\sum_{n-2m+3}^{n-m+1}\mu_i(\phi)\right)\\&=\frac{n-m+1}{n-k+1}\mu(\phi)\\&\quad+\frac{m-k}{n-k+1}\left(\sum_{i=1}^{m-1}\mu_i(\phi)+\sum_{n-2m+3}^{n-m+1}\mu_i(\phi)\right)\\&\leqslant \mu(\phi)+\frac{(m-k)(2m-2)}{n-k+1}.\end{aligned}$$

On the other hand, a lower bound may be obtained by assuming a maximal overcounting factor of $m-k+1$ for all occurrences of ϕ, regardless of position within ω. This time,

$$\begin{aligned}\mathrm{fr}_{\omega}^{k}(\phi) &\geqslant \frac{1}{n-k+1}\sum_{i=1}^{n-m+1}\mu_i(\phi)\\&=\frac{n-m+1}{n-k+1}\mu(\phi)\\&\geqslant \mu(\phi)-\frac{m-k}{n-k+1}.\end{aligned}$$

We observe that

$$0\leqslant\frac{m-k}{n-k+1}\leqslant\frac{(2m-2)(m-k)}{n-k+1}.$$

Thus, if we define

$$S(k,m,n)=\frac{(2m-2)(m-k)}{n-k+1}, \tag{5}$$

then

$$\left|\mathrm{fr}_{\omega}^{k}(\phi)-\mu(\phi)\right|\leqslant S(k,m,n).$$

Let us now define

$$N_\epsilon=\frac{(2m-2)(m-k)}{\epsilon}+k-1.$$

Then for all $n>\max\{N_\epsilon,3m-2\}$ and all $\omega\in N_m(\Gamma_\epsilon)$, $|\omega|=n$, we also have $\mathrm{fr}_{\omega}^{k}\in\Gamma$, i.e., $\omega\in\mathcal{B}(\Gamma)$. Hence, $N_m(\Gamma_\epsilon)\subseteq^e\mathcal{B}(\Gamma)$ as claimed. ■

Note that unlike Construction A, here we obtain that $N_m(\Gamma_\epsilon)\subseteq^e\mathcal{B}(\Gamma)$ for every $m\geqslant k$. We also note $N_\epsilon=\Omega(\frac{1}{\epsilon})$.

A stronger statement than that of Theorem 14 may be made in the case of WSCSs, in which ϵ is removed. This is due to the fact that the quantity $S(k,m,n)$ defined by equation (5) in the proof of Theorem 14 is in fact $o(1)$ for constant k and m.

Corollary 15: Let Γ be a convex fat SCS, fix $m\geqslant k$, and define the tolerance function $\xi(n)=S(k,m,n)$, where $S(k,m,n)$ is defined in (5)*. Then*

$$N_m(\Gamma)\subseteq^e\overline{\mathcal{B}}(\Gamma).$$

Proof: The proof follows from the proof of Theorem 14, by noting that $S(k,m,n)$ is both $o(1)$ and $\Omega(\frac{1}{n})$. ■

Theorem 16: Let Γ be a closed convex fat SCS. Then

$$\limsup_{m\to\infty}\mathsf{cap}(N_m(\Gamma))=\mathsf{cap}(\mathcal{B}(\Gamma)).$$

Proof: By Corollary 15 and Theorem 8

$$\mathsf{cap}(N_m(\Gamma)))\leqslant\mathsf{cap}(\overline{\mathcal{B}}(\Gamma))=\mathsf{cap}(\mathcal{B}(\Gamma)).$$

Note that this statement does not require taking m to infinity, and it applies to all $m\geqslant k$.

For the other direction, we contend that for every Γ-feasible $\epsilon>0$ there exists M_ϵ such that for all $m>M_\epsilon$

$$\mathsf{cap}\left(N_{m^2}(\Gamma)\right)\geqslant\mathsf{cap}(R_m(\Gamma_\epsilon)).$$

To prove this claim, let $\omega=\omega_1\omega_2\dots\omega_\ell\in R_m(\Gamma_\epsilon)$ with $\ell\geqslant m$ and $\omega_i\in\mathcal{B}_m(\Gamma_\epsilon)$ for all i. Denote $\mu_i=\mathrm{fr}_{\omega_i}^{k}\in\Gamma_\epsilon$. Let ω' be any length-m^2 subword of ω, and let us check the frequency any k-tuple $\phi\in\Sigma^k$ appears in it.

Such a sequence ω' is surely fully contained in some $m+1$ consecutive subwords, say $\omega_j\omega_{j+1}\dots\omega_{j+m}$. Let us denote

$$\mu=\frac{1}{m+1}\sum_{i=0}^{m}\mu_{j+i}.$$

Again, $\mu\in\Gamma_\epsilon$ due to the convexity of Γ_ϵ.

In a similar fashion to previous proofs, the frequency of ϕ in ω' is easily seen to be upper bounded by

$$\begin{aligned}\mathrm{fr}_{\omega'}^{k}(\phi)&\leqslant\frac{(m-k+1)\sum_{i=0}^{m}\mu_{j+i}(\phi)+(k-1)m}{m^2-k+1}\\&\leqslant\mu(\phi)+\frac{m}{m^2-k+1},\end{aligned}$$

by accounting for the occurrences of ϕ in each subword ω_{j+i}, and upper bounding the effect of the m concatenation points between those subwords.

Conversely, the $m-1$ subwords $\omega_{j+1}\omega_{j+2}\dots\omega_{j+m-1}$ must be fully contained within ω'. Thus, we obtain the lower bound

$$\begin{aligned}\mathrm{fr}_{\omega'}^{k}(\phi)&\geqslant\frac{(m-k+1)\sum_{i=1}^{m-1}\mu_i(\phi)}{m^2-k+1}\\&\geqslant\mu(\phi)-\frac{mk+2(m-k+1)}{m^2-k+1}.\end{aligned}$$

It therefore follows that

$$\left|\mathrm{fr}_{\omega'}^{k}(\phi)-\mu(\phi)\right|\leqslant\frac{mk+2(m-k+1)}{m^2-k+1}. \tag{6}$$

Since the right-hand side of (6) is $o(1)$ as $m\to\infty$, there exists $M_\epsilon\geqslant k$ such that for all $m>M_\epsilon$, $\mathrm{fr}_{\omega'}^{k}\in\Gamma$, and thus

$$R_m(\Gamma_\epsilon)\subseteq N_{m^2}(\Gamma).$$

as claimed. It follows that for $m>M_\epsilon$,

$$\mathsf{cap}\left(N_{m^2}(\Gamma)\right)\geqslant\mathsf{cap}(R_m(\Gamma_\epsilon)).$$

Taking $\limsup_{m\to\infty}$ on both sides we obtain

$$\begin{aligned}\limsup_{m\to\infty}\mathsf{cap}\left(N_m(\Gamma)\right)&\geqslant\limsup_{m\to\infty}\mathsf{cap}\left(N_{m^2}(\Gamma)\right)\\&\geqslant\limsup_{m\to\infty}\mathsf{cap}(R_m(\Gamma_\epsilon))\\&=\mathsf{cap}(\mathcal{B}(\Gamma_\epsilon)),\end{aligned}$$

where the last equality is due to Theorem 12. Now, since this holds for all Γ-feasible $\epsilon>0$, taking the limit as $\epsilon\to 0$, by the continuity guaranteed in Theorem 8 we get

$$\limsup_{m\to\infty}\mathsf{cap}\left(N_m(\Gamma)\right)\geqslant\mathsf{cap}(\mathcal{B}(\Gamma)),$$

which completes the proof. ■

The graph G that represents $N_m(\Gamma)$, is deterministic, and therefore has anticipation 0. The graph is not necessarily an encoder (due to an unequal out-degree), but by using the state-splitting algorithm on G we may generate a $p : q$ encoder, at the cost of possibly increasing the anticipation.

Apart from describing two constructions for encoders, we have thus far also proved the following corollary.

Corollary 17: Let Γ be a closed convex fat SCS. Then

$$\mathsf{cap}^{\subseteq}(\Gamma) = \mathsf{cap}(\mathcal{B}(\Gamma)).$$

Proof: This is immediate either from Corollary 13 or from Theorem 16. ∎

C. A Short Case Study

As a short case study we provide the following example. Consider the SCS over $\Sigma = \{0, 1\}$, which is defined by the set

$$\Gamma = \left\{\mu \in \mathcal{P}(\Sigma^2) \ : \ \mu(11) \leqslant 0.205\right\}.$$

This SCS was called the $(0, 1, 0.205)$-RLL SCS in [5], and its capacity is $\mathsf{cap}(\mathcal{B}(\Gamma)) \approx 0.98$. We investigate the encoders presented thus far, with an intention of building an encoder with rate $\frac{3}{4}$.

We first focus on the block encoder associated with $R_m(\Gamma)$. Choosing $\epsilon = 0.005$, a quick use of (4) shows that any $m > 200$ guarantees that we satisfy the semiconstraints. A finer analysis, accounting for divisibility conditions, reveals all $m > 156$ suffice. The latter is indeed tight, since for $m = 156$ we have $\omega = 1^{32}0^{123}1 \in \mathcal{B}_{156}(\Gamma_{0.005})$, but

$$\lim_{i\to\infty} \mathrm{fr}^2_{\omega^i}(11) = \frac{32}{156} > 0.205,$$

so for large enough i, ω^i does not satisfy the semiconstraints. However, there exist smaller values of m which are acceptable. The smallest one is $m = 5$. However, in this case, $|\mathcal{B}_5(\Gamma_{0.005})| = 13$, not achieving the required rate of $\frac{3}{4}$. The next possible acceptable value is $m = 10$, in which case $|\mathcal{B}_{10}(\Gamma_{0.005})| = 379$, exceeding the required rate, but at the cost of having an unwieldy number of edges in the encoder.

On the other hand, the encoder associated with $N_m(\Gamma)$ is simpler. We can choose $m = 6$. We first construct the De-Bruijn graph of order $m - 1 = 5$ in which we eliminate vertices that correspond to a word with more than a single appearance of the pattern 11. Since we would like to build an encoder with rate $\frac{3}{4}$, we take the graph to its 4th power, and keep the appropriate irreducible subgraph. After combining vertices with the same follower sets and applying the state-splitting algorithm, we obtain an encoder with 14 vertices and 112 edges. Three rounds of state splitting were used, thus the anticipation as at most 3.

IV. Approaching Capacity From Above

In this section we consider the dual question to the one asked in Section III: which fully constrained systems, represented as the language of a directed labeled graph, *contain* a given semiconstrained system. Additionally, we would like to know whether the capacity of a sequence of those fully constrained system can approach the capacity of the semiconstrained system in the limit. In an analogous fashion to Definition 9, we define the following.

Definition 18: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a SCS. We define

$$\mathsf{cap}^{\supseteq}(\Gamma) = \inf_{\mathcal{L}(G)\supseteq\mathcal{B}(\Gamma)} \mathsf{cap}(\mathcal{L}(G)). \tag{7}$$

As we shall soon see, the result is quite pessimistic. We first give an auxiliary lemma, and then proceed to prove the main theorem. For this lemma we require the following definition.

Definition 19: Let $\eta \in \mathcal{P}(\Sigma^k)$ be a rational measure, i.e., for every $a \in \Sigma^k$, $\eta(a) \in \mathbb{Q}$. We define the following graphs nG_η, for each $n \in \mathbb{N}$. Let $M \in \mathbb{N}$ be the smallest natural number such that $M\eta$ is an integer vector. The initial vertex set of nG_η is $V = \Sigma^{k-1}$. For each $a_0, a_1, \dots, a_{k-1} \in \Sigma$, we place $nM\eta(a_0\ a_1 \dots a_{k-1})$ parallel edges from vertex $a_0\ a_1 \dots a_{k-2}$ to vertex $a_1\ a_2 \dots a_{k-1}$. Finally, we remove vertices with zero in-degree and out-degree, i.e., isolated vertices.

Lemma 20: Let $\eta \in \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ be a positive (entry-wise) rational and shift-invariant measure. Then for any $\alpha \in \Sigma^*$ there exists $\beta \in \Sigma^*$ such that $\mathrm{fr}^k_{\alpha\beta} = \eta$.

Proof: Since η is rational and shift invariant, for all $a_0, a_1, \dots, a_{k-1} \in \Sigma$ we have that $\eta(a_0\ a_1 \dots a_{k-1}) \in \mathbb{Q}$, and

$$\sum_{b\in\Sigma} \eta(a_0\ a_1 \dots a_{k-2}b) = \sum_{b\in\Sigma} \eta(ba_0\ a_1 \dots a_{k-2}).$$

Assume we are given a sequence $\alpha \in \Sigma^m$ with $m \geqslant k$ (if $m < k$ we arbitrarily extend α so its length is at least k).

We now consider the graph $(m + 1)G_\eta$. We note that since η is positive, the graph $(m + 1)G_\eta$ is strongly connected, i.e., there is a directed path between any source vertex and destination vertex. Additionally, the shift-invariance property of η implies that the in-degree of every vertex equals its out-degree.

With a directed path of n edges in the graph we associate a sequence of length $n + k - 1$ over Σ via a sliding-window reading of the sequence. Formally, a sequence $\alpha = a_0\ a_1 \dots a_{n+k-2} \in \Sigma^{n+k-1}$, $a_i \in \Sigma$, is associated with the directed path whose ith edge is $a_i a_{i+1} \dots a_{i+k-2} \to a_{i+1}a_{i+2} \dots a_{i+k-1}$, for all $0 \leqslant i \leqslant n - 1$. Since this mapping is a bijection (up to parallel edges), by abuse of notation we shall refer to α as both the sequence and the path.

The given sequence α describes a path in $(m+1)G_\eta$, where the graph parameter $(m + 1)$ ensures the path can consist of distinct (though perhaps parallel) edges. Let us remove the edges of this path from the graph to obtain a graph G'.

First we note that G' is still strongly connected since any two vertices originally connected by an edge (perhaps several parallel edges) are still connected by at least one edge. This is because a total of $m-k+1 < m+1$ edges were removed, and two vertices connected by an edge in $(m + 1)G_\eta$ are actually connected by at least $m + 1$ edges.

Next we distinguish between two cases. If the removed path α was a cycle, then G' still has the property that every vertex has an equal in-degree and out-degree. Thus, it contains an Eulerian cycle (going over every edge exactly once) which we denote as β. Since the ending vertex of α still has a positive out-degree, we may, without loss of generality, choose β to begin in the same vertex. It follows that $\alpha\beta$ is an Eulerian cycle in the original graph $(m+1)G_\eta$.

In the second case, the removed path α is not a cycle. In that case, except for the starting vertex and ending vertex of α, all other vertices have equal in-degree and out-degree. The starting vertex of α has an in-degree larger by 1 compared with its out-degree, and the ending vertex of α has the reversed situation. Thus, G' contains an Eulerian path β that starts in the ending vertex of α, and ends in the starting vertex of α. Again, $\alpha\beta$ is therefore an Eulerian cycle in the original graph $(m+1)G_\eta$.

We note that the sequence associated with the sliding-window reading induced by the path $\alpha\beta$ in $(m+1)G_\eta$ has each window $\phi \in \Sigma^k$ appear exactly as an $\eta(\phi)$ fraction of the windows of size k, i.e., $\mathrm{fr}^k_{\alpha\beta}(\phi) = \eta(\phi)$. ■

Corollary 21: Let Γ be a fat SCS. Then for all $\alpha \in \Sigma^*$ there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in \mathcal{B}(\Gamma)$, i.e., any finite prefix may be completed to a word in the semiconstrained system.

Proof: Since Γ is fat, there exists a probability measure $\nu \in \mathrm{int}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k))$, i.e., ν is shift invariant and in the interior of Γ. We can take a sequence of rational shift-invariant probability measures that converge to ν, and since there exists an $\epsilon > 0$ environment of ν contained within $\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k)$, then we deduce the existence of a strictly positive and rational shift-invariant probability measure $\eta \in \Gamma$. From Lemma 20 we can find $\beta \in \Sigma^*$ such that $\alpha\beta \in \mathcal{B}(\Gamma)$. ■

We note that the property that every prefix may be extended to a word in $\mathcal{B}(\Gamma)$, is called *right density* in formal-language theory (see [18]).

We now state the main result of this section.

Theorem 22: Let Γ be a fat SCS. Let Γ' be a fully constrained system such that $\mathcal{B}(\Gamma) \subseteq \mathcal{B}(\Gamma')$. Then $\mathcal{B}(\Gamma') = \Sigma^*$, and thus

$$\mathsf{cap}^{\supseteq}(\Gamma) = \log_2 |\Sigma|.$$

Proof: Consider any $\alpha \in \Sigma^*$. By Corollary 21, there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in \mathcal{B}(\Gamma)$, and hence, $\alpha\beta \in \mathcal{B}(\Gamma')$ as well. Since Γ' is factorial, $\alpha \in \mathcal{B}(\Gamma')$. Hence, $\mathcal{B}(\Gamma') = \Sigma^*$. ■

We note the peculiar asymmetry between fully constrained systems *contained* within a fat SCS, and fully constrained systems *containing* a fat SCS. While in the former we have a sequence of such fully constrained systems that approach the capacity of the given fat SCS, in the latter there is exactly one fully constrained system containing the SCS, and that is the entire space Σ^*.

V. Combining SCSs With Combinatorial Constraints

In our discussion thus far, we required any SCS Γ to be fat. Unfortunately, if for some $\phi \in \Sigma^k$ we have $\mu(\phi) = 0$ for all $\mu \in \Gamma$, then Γ is not fat. Intuitively, in that case Γ does not occupy all the dimensions of Σ^k. It follows that none of the results obtained in the previous sections apply to fully constrained systems, since the latter employ such zero constraints, and are therefore not fat SCSs. In the following we discuss how this situation can be resolved by defining relatively-fat SCSs.

We start by defining the set of forbidden words, that is, those words which are never subwords of the admissible words of the given SCS.

Definition 23: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be a set of probability measures. We denote by $\mathcal{F}(\Gamma) \subseteq \Sigma^k$ the following set of Γ-forbidden *k-tuples,*

$$\mathcal{F}(\Gamma) = \left\{ \phi \in \Sigma^k \ : \ \forall \mu \in \Gamma, \mu(\phi) = 0 \right\}.$$

We now define a relatively fat set Γ. Informally, we call a set Γ *relatively fat (RF)* if apart from $\mathcal{F}(\Gamma)$ it is fat. A formal definition follows.

Definition 24: For any $\Gamma \subseteq \mathcal{P}(\Sigma^k)$, let us denote $\mathcal{D} = \Sigma^k \setminus \mathcal{F}(\Gamma)$. We say that Γ is relatively fat (RF) *if*

$$\mathrm{cl}_{\mathcal{D}}(\mathrm{int}_{\mathcal{D}}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\mathcal{D}))) = \mathrm{cl}_{\mathcal{D}}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\mathcal{D})),$$

where $\mathrm{cl}_{\mathcal{D}}$ and $\mathrm{int}_{\mathcal{D}}$ are the closure and interior with respect to $\mathcal{D}$, respectively, and $\mathcal{P}_{\mathrm{si}}(\mathcal{D})$ denotes the set of shift-invariant measures on $\mathcal{D}$.

We mention briefly that other definitions of RF sets are possible, somewhat generalizing the definition we use. For example, one may generalize the definition to one that requires Γ to be contained within an affine space of dimension possibly lower than $\mathcal{P}(\Sigma^k)$, and further that Γ is fat with respect to that affine space. However, such generality is not required by us.

We now go through the results obtained thus far for fat SCSs, and describe the necessary changes required to make them work for RF SCSs as well. We first note that the general form of Theorem 8, which holds for every set Γ and is given in [4, Ch. 3] is as follows (with modified notations).

Theorem 25: Let $\Gamma \subseteq \mathcal{P}(\Sigma^k)$ be closed and convex. Then

$$\begin{aligned}\log_2 |\Sigma| - \inf_{\eta \in \mathrm{int}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k))} D(\eta \| \eta') &\leqslant \mathsf{cap}(\mathcal{B}(\Gamma)) \\ &\leqslant \log_2 |\Sigma| - \inf_{\eta \in \mathrm{cl}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k))} D(\eta \| \eta'),\end{aligned}$$

where int *and* cl *are the interior and closure of a set.*

In the case of RF SCSs, the interior of Γ may be empty and therefore Theorem 25 states that the capacity of a RF SCS is bounded from below by $-\infty$. We are therefore left with the upper bound of

$$\mathsf{cap}(\mathcal{B}(\Gamma)) \leqslant \log_2 |\Sigma| - \inf_{\eta \in \mathrm{cl}(\Gamma \cap \mathcal{P}_{\mathrm{si}}(\Sigma^k))} D(\eta \| \eta').$$

Construction A does not work for RF SCSs. The cause of failure is the obvious concatenation point between blocks, which may contain a forbidden word. For example, assume a binary alphabet, $k = 2$, and $\mathcal{F} = \{11\}$, i.e., the SCS is in fact the $(1, \infty)$-RLL fully constrained system. In this case, taking two words of length m, the first, α_1, ending with a 1, and the second α_2, starting with a 1, and concatenating them together, will create the forbidden pattern 11 in $\alpha_1\alpha_2$. A way of solving this problem is by placing a carefully crafted string, β, between the two blocks, i.e., $\alpha_1\beta\alpha_2$. As long $|\beta| = o(m)$, Construction A works. A similar method, developed for fully constrained multidimensional RLL systems was described in [6].

In comparison, Construction B indeed does work for RF SCSs. The only change needed in the proofs is to alter the definition of Γ_ϵ from (1) to

$$\Gamma_\epsilon = \Big\{\eta \in \Gamma : \forall \phi \in \Sigma^k \setminus \mathcal{F}(\Gamma), \inf_{\mu \in \mathcal{P}(\Sigma^k)\setminus\Gamma} |\eta(\phi) - \mu(\phi)| > \epsilon\Big\}.$$

Having considered the generalization of Section III to RF SCSs, we now turn to discuss the generalization of Section IV. In what follows, we do not even require the relative-fatness property. We are therefore interested in fully constrained systems containing a given SCS. Unlike contained fully constrained system, in the containing case the discussion is somewhat more involved.

We recall some useful notions from graph theory. Two vertices, v_1 and v_2, in a directed graph G, are said to be *bi-connected* if there is a directed path from v_1 to v_2, and a directed path from v_2 to v_1. Bi-connectedness is an equivalence relation, and its equivalence classes are called *strongly connected components*.

In Corollary 21 we used the fact that for a fat Γ, there exists a rational $\eta \in \Gamma$ such that G_η is a single strongly connected component. Unfortunately, this is no longer the case for general SCSs, even if we restrict ourselves to RF SCSs as shown in the following example.

Example 26: Fix $\Sigma = \{0, 1\}$, and $k = 4$. Define $\mu_1, \mu_2 \in \mathcal{P}_{\text{si}}(\Sigma^k)$ as follows:

$$\mu_1 = \delta_{1111}, \quad \text{and} \quad \mu_2 = \frac{1}{2}(\delta_{1010} + \delta_{0101}),$$

where δ_ϕ denotes probability measure of value 1 at ϕ, and 0 elsewhere. Let Γ be the convex hull of μ_1 and μ_2. Then Γ is non-empty, convex, relatively fat, and contains only shift-invariant measures. However, except for μ_1 and μ_2, there is no other $\eta \in \Gamma$ such that G_η has a single strongly connected component. □

The following definition is intended to capture and isolate this kind of pathological behaviour.

Definition 27: Let $\Gamma \subseteq \mathcal{P}_{\text{si}}(\Sigma^k)$ be a SCS. The essential part *of Γ is defined as*

$$\text{ess}(\Gamma) = \{\eta \in \Gamma \ : \ \mathcal{B}(\{\eta\}) \neq \emptyset\}.$$

Thus, ess(Γ) keeps only those measures of Γ that have at least one admissible word. Note that by definition, ess(Γ) contains only rational measures. We also note that even if Γ is convex, the set ess(Γ) may not necessarily be convex (even if we consider only convex rational combinations of measures from ess(Γ)). This can be seen in Example 26, in which $\text{ess}(\Gamma) = \{\mu_1, \mu_2\}$.

Lemma 28: Let $\eta \in \mathcal{P}_{\text{si}}(\Sigma^k)$ be a rational shift-invariant measure. Then $\mathcal{B}(\{\eta\}) \neq \emptyset$ if and only if G_η is strongly connected after removing isolated vertices.

Proof: In the first direction assume $\mathcal{B}(\{\eta\}) \neq \emptyset$. Then let $\omega \in \mathcal{B}(\{\eta\})$. Following the same steps as in the proof of Lemma 20, we obtain that ω corresponds to an Eulerian cycle in nG_η for some $n \in \mathbb{N}$ (after removing isolated vertices). Thus, nG_η is strongly connected, and since n does not affect this property, G_η is also strongly connected.

In the other direction, assume G_η is strongly connected. Since η is shift invariant, the in-degree and out-degree of each vertex are equal, and there exists an Eulerian cycle in G_η. Again, by the proof of Lemma 20, this cycle corresponds to a word $\omega \in \Sigma^*$ with $\text{fr}^k_\omega = \eta$. Thus, $\omega \in \mathcal{B}(\{\eta\})$. ■

The next step we take is to define the essential graph of a SCS.

Definition 29: Let $\Gamma \subseteq \mathcal{P}_{\text{si}}(\Sigma^k)$. Denote by $G_{\text{ess}}(\Gamma)$ the following directed labeled graph: Vertices are represented by elements of Σ^{k-1}. For each $\phi = a_0\, a_1 \dots a_{k-1} \in \Sigma^k$, such that there exists some $\eta \in \text{ess}(\Gamma)$ with $\eta(\phi) > 0$, we place an edge $a_0 a_1 \dots a_{k-2} \to a_1\ a_2 \dots a_{k-1}$, labeled by a_0. Any isolated vertices (i.e., vertices with both in-degree and out-degree of zero) are then removed.

Note that the edge $a_0 a_1 \dots a_{k-2} \to a_1\ a_2 \dots a_{k-1}$ is labeled by a_0 and not by a_{k-1}. This simplifies the notation later.

Intuitively, $G_{\text{ess}}(\Gamma)$ is the union of all G_η, $\eta \in \text{ess}(\Gamma)$, where parallel edges are merged, and isolated vertices are removed. Here, we define the union of two directed labeled graphs, $G_1 = (V_1, E_1)$, and $G_2 = (V_2, E_2)$, with $E_i \subseteq V_i \times V_i \times \Sigma$, as $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$. We note that the sets of vertices of the two original graphs are not necessarily disjoint, and the same goes for the sets of edges. We collect some more insight into words having a shift-invariant measure.

Lemma 30: Let $\eta \in \mathcal{P}_{\text{si}}(\Sigma^k)$ be a shift-invariant measure, and let $\omega = a_0 a_1 \dots a_{n-1} \in \Sigma^n$ be a word for which $\text{fr}^k_\omega = \eta$. Then $a_0 \dots a_{k-2} = a_{n-k+1} \dots a_{n-1}$, i.e., the $(k-1)$-prefix and $(k-1)$-suffix of ω are equal.

Proof: Let ω be a word with k-tuple distribution given by η. Since η is shift invariant, for every $\phi \in \Sigma^{k-1}$ we have

$$\sum_{a\in\Sigma} \eta(a\phi) = \sum_{a\in\Sigma} \eta(\phi a). \tag{8}$$

In particular, let us examine $\phi = a_0\ a_1 \dots a_{k-2}$, the $(k-1)$-prefix of ω. The left-hand side of (8) is given by

$$\frac{|\{(\alpha,\gamma) \ : \ \alpha,\gamma \in \Sigma^*, \alpha\phi\gamma = \omega\}| - 1}{n-k+2}, \tag{9}$$

where the subtraction of 1 in the numerator is due to the fact we require a letter to appear before ϕ, and therefore cannot count its appearance as a prefix of ω.

Assume to the contrary that ϕ is not the $(k-1)$-suffix of ω. In that case, every occurrence of ϕ in ω is followed by a letter, and then the right-hand side of (8) is

$$\frac{|\{(\alpha,\gamma) \ : \ \alpha,\gamma \in \Sigma^*, \alpha\phi\gamma = \omega\}|}{n-k+2},$$

but that differs from (9), a contradiction. ■

We shall call a word $\omega \in \Sigma^*$ *k-shift-invariant* if $\text{fr}^k_\omega \in \mathcal{P}_{\text{si}}(\Sigma^k)$. By the previous lemma, the $(k-1)$-suffix of ω equals its $(k-1)$-prefix. We shall therefore find it convenient to chop off the $(k-1)$-suffix of ω using the following operator. If $\omega = a_0\ a_1 \dots a_{n-1}$, $n \geqslant k-1$, then we define

$$\text{SuffChop}_{k-1}(\omega) = a_0\ a_1 \dots a_{n-k}.$$

The following corollary is therefore immediate.

Lemma 31: Let $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ be a SCS. Then for every $\omega \in \mathcal{B}(\Gamma)$ there exists a cycle in $G_{\mathrm{ess}}(\Gamma)$ generating $\mathrm{SuffChop}_{k-1}(\omega)$. *Hence,*

$$\mathcal{B}(\Gamma) \subseteq \mathcal{L}(G_{\mathrm{ess}}(\Gamma)).$$

Proof: Let $\omega \in \mathcal{B}(\Gamma)$, and denote $|\omega| = n$. We can assume $n \geqslant k-1$. Additionally, we must have $\eta = \mathrm{fr}^k_\omega \in \mathrm{ess}(\Gamma)$, by definition. If we read ω by a sliding window of size $k-1$, then by Lemma 30 we get a sequence of vertices of $G_{\mathrm{ess}}(\Gamma)$ forming a cycle. The labels along this cycle generate $\mathrm{SuffChop}_{k-1}(\omega)$. We can then take again the first $k-1$ edges of the cycle to complete a reading of ω. Thus, $\omega \in \mathcal{L}(G_{\mathrm{ess}}(\Gamma))$. ■

We argue that every word obtained by reading the labels of edges along a walk on $G_{\mathrm{ess}}(\Gamma)$ can be completed to a word in $\mathcal{B}(\Gamma)$.

Theorem 32: Let $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ be a convex SCS. Then, for every $\alpha \in \mathcal{L}(G_{\mathrm{ess}}(\Gamma))$ there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in \mathcal{B}(\Gamma)$.

Proof: Let $\alpha \in \mathcal{L}(G_{\mathrm{ess}}(\Gamma))$, $|\alpha| = n$, be a word that is obtained by reading the labels of edges $e_1 \to e_2 \to \cdots \to e_n$ along a path in $G_{\mathrm{ess}}(\Gamma)$.

Each edge e_i corresponds to some $\phi_i \in \Sigma^k$. By the definition of $G_{\mathrm{ess}}(\Gamma)$, the edge e_i exists since there exists $\eta_i \in \Gamma$, such that $\eta_i(\phi_i) > 0$ and $\mathcal{B}(\eta_i) \neq \emptyset$. Thus, η_i is rational and shift invariant. By Lemma 28, G_{η_i} is strongly connected (after removing isolated vertices).[1]

We now take a convex combination

$$\eta = \sum c_i \eta_i,$$

where $c_i > 0$, $c_i \in \mathbb{Q}$, for all $1 \leqslant i \leqslant n$, and $\sum_{i=1}^n c_i = 1$. Since Γ is convex, we have $\eta \in \Gamma$. By our previous observations, G_η contains the path $e_1 \to \cdots \to e_n$, and is the union of the graphs $\left\{G_{\eta_i}\right\}_{i=1}^n$, each of which is strongly connected. Thus, G_η is also strongly connected.

We first note that by Lemma 28, $\mathcal{B}(\{\eta\}) \neq \emptyset$, i.e., $\eta \in \mathrm{ess}(\Gamma)$. Following the same reasoning as in the proof of Lemma 20, there exists $m \in \mathbb{N}$ such that there exists an Eulerian cycle in mG_η starting with the path $e_1 \to \cdots \to e_n$. This Eulerian cycle therefore corresponds to a reading of a word $\alpha\beta$, by sliding windows of size k, for which $\mathrm{fr}^k_{\alpha\beta} = \eta \in \Gamma$. Hence, $\alpha\beta \in \mathcal{B}(\Gamma)$. ■

As a corollary we obtain the main result of this section.

Corollary 33: Let $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ be a convex SCS. Then $\mathcal{L}(G_{\mathrm{ess}}(\Gamma))$ is the unique smallest fully constrained system containing $\mathcal{B}(\Gamma)$. In particular,

$$\mathsf{cap}^{\supseteq}(\Gamma) = \mathsf{cap}(\mathcal{L}(G_{\mathrm{ess}}(\Gamma))).$$

Proof: Let G be a directed graph, with edge labels from Σ, such that $\mathcal{B}(\Gamma) \subseteq \mathcal{L}(G)$. Consider a word $\alpha \in \mathcal{L}(G_{\mathrm{ess}}(\Gamma))$. By Theorem 32 there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in \mathcal{B}(\Gamma)$. Thus, $\alpha\beta \in \mathcal{L}(G)$. In particular, a prefix of a path generating $\alpha\beta$ in G, generates α. Hence, $\alpha \in \mathcal{L}(G)$, and $\mathcal{L}(G_{\mathrm{ess}}(\Gamma)) \subseteq \mathcal{L}(G)$. The claims now follow. ■

We devote the remainder of the section for some curious observations. Our first observation is that while one might initially assume $\mathsf{cap}^{\supseteq}(\Gamma)$ is monotone increasing in the capacity, this is not generally the case, as the following example shows.

Example 34: Let Γ_1, Γ_2 be SCSs over $\Sigma = \{0,1\}$ with $k = 3$ defined by

$$\Gamma_1 = \left\{\mu \in \mathcal{P}_{\mathrm{si}}(\Sigma^k) \ : \ \mu(000), \mu(111), \mu(101) \leqslant 0.01\right\}.$$

$$\Gamma_2 = \left\{\mu \in \mathcal{P}_{\mathrm{si}}(\Sigma^k) \ : \ \mu(000) = 0\right\}.$$

We note that both Γ_1 and Γ_2 are shift invariant, convex, and relatively fat. The capacity of Γ_1 may be obtained using Theorem 8, and that of Γ_2 is also easily obtained since it is also a fully constrained system. We reach

$$\mathsf{cap}(\mathcal{B}(\Gamma_1)) \approx 0.462, \qquad \mathsf{cap}(\mathcal{B}(\Gamma_2)) \approx 0.879.$$

One can easily see that $G_{\mathrm{ess}}(\Gamma_1)$ is the De-Bruijn graph of order 2. Thus,

$$\mathsf{cap}^{\supseteq}(\Gamma_1) = 1.$$

Since Γ_2 is fully constrained to begin with,

$$\mathsf{cap}^{\supseteq}(\Gamma_2) = \mathsf{cap}(\mathcal{B}(\Gamma_2)) \approx 0.879.$$

Hence,

$$\mathsf{cap}(\mathcal{B}(\Gamma_1)) < \mathsf{cap}(\mathcal{B}(\Gamma_2)),$$

but

$$\mathsf{cap}^{\supseteq}(\Gamma_1) > \mathsf{cap}^{\supseteq}(\Gamma_2).$$

□

While a gap may exist between $\mathsf{cap}(\mathcal{B}(\Gamma))$ and $\mathsf{cap}^{\supseteq}(\Gamma)$, as is demonstrated in Example 34, this is never the case with zero capacity.

Lemma 35: Let $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ be a convex SCS. Then we have $\mathsf{cap}(\mathcal{B}(\Gamma)) = 0$ if and only if $\mathsf{cap}^{\supseteq}(\Gamma) = 0$.

Proof: For the first direction, assume we have $\mathsf{cap}^{\supseteq}(\Gamma) = 0$. By definition,

$$\mathsf{cap}(\mathcal{B}(\Gamma)) \leqslant \mathsf{cap}^{\supseteq}(\Gamma) = 0,$$

and therefore $\mathsf{cap}(\mathcal{B}(\Gamma))$ is either 0 or $-\infty$. We note that $\mathcal{B}(\Gamma) \neq \emptyset$, since otherwise we would have $\mathcal{L}(G_{\mathrm{ess}}(\Gamma)) = \emptyset$ implying $\mathsf{cap}^{\supseteq}(\Gamma) = -\infty$, a contradiction. It follows that we must have some $\omega \in \mathcal{B}(\Gamma)$, $|\omega| \geqslant k$. Denote $\omega = \alpha\beta$, where $\alpha = \mathrm{SuffChop}_{k-1}(\omega)$. But then

$$\mathrm{fr}^k_{\alpha^n\beta} = \mathrm{fr}^k_{\alpha\beta} \in \Gamma,$$

for all $n \in \mathbb{N}$, and $\mathcal{B}(\Gamma)$ is an infinite set, giving us the desired $\mathsf{cap}(\mathcal{B}(\Gamma)) = 0$.

In the other direction, assume $\mathsf{cap}(\mathcal{B}(\Gamma)) = 0$. By Lemma 31, $\mathcal{B}(\Gamma) \subseteq \mathcal{L}(G_{\mathrm{ess}}(\Gamma))$. Let $\omega \in \mathcal{B}(\Gamma)$, with $\mathrm{fr}^k_\omega = \eta$. Then G_η (after removing isolated vertices) is Eulerian, since η is shift invariant. We note that this Eulerian cycle must be simple, otherwise nG_η contains an exponential (in n) number of Eulerian cycles, implying $\mathsf{cap}(\mathcal{B}(\Gamma)) > 0$, a contradiction.

Now, assume $\omega_1, \omega_2 \in \mathcal{B}(\Gamma)$ be two distinct words, with $\mathrm{fr}^k_{\omega_i} = \eta_i$. As in the proof of Theorem 32, the convexity of Γ implies that the simple Eulerian cycles of G_{η_1} and G_{η_2} are either identical or disjoint. Otherwise, an appropriate rational convex combination of η_1 and η_2 results in some $\eta \in \Gamma$ whose

[1] Throughout the proof we remove isolated vertices from graphs.

G_η is Eulerian and non-simple, implying a positive capacity for $\mathcal{B}(\Gamma)$, a contradiction.

It follows that $G_{\mathrm{ess}}(\Gamma)$ is a union of disjoint simple Eulerian cycles. Thus, $\mathsf{cap}^{\supseteq}(\Gamma) = \mathsf{cap}(\mathcal{L}(G_{\mathrm{ess}}(\Gamma))) = 0$. ■

By the proof of Lemma 35 we also observe that for a convex $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$ with $\mathsf{cap}(\mathcal{B}(\Gamma)) = 0$, we have that $\mathcal{B}(\Gamma)$ is a regular language, though not necessarily a fully constrained system. This is no longer true if we omit the requirement that $\Gamma \subseteq \mathcal{P}_{\mathrm{si}}(\Sigma^k)$, as the following example shows.

Example 36: Let $\Sigma = \{0, 1\}$, $k = 2$, *and define*

$$\Gamma = \left\{ \mu \in \Sigma^2 \ : \ \mu(00) < \frac{1}{2}, \mu(11) < \frac{1}{2}, \mu(01) = 0 \right\}.$$

We note that Γ *is relatively fat, convex, but contains some measures which are not shift invariant. Interestingly,*

$$\mathcal{B}(\Gamma) = \left\{ 1^n 0^n \ : \ n \in \mathbb{N} \right\},$$

which is not a regular language. □

VI. Conclusions and Discussion

This work was devoted to fully constrained systems either contained or containing a given SCS. This is in order to find connections between SCSs and fully constrained systems, further motivated by the extensive literature on encoders for fully constrained systems. Apart from two encoder constructions, an interesting asymmetry between contained and containing fully constrained systems emerged. Whereas the former approach the capacity of the given SCS, the latter are generally bounded away from it.

We suspect cleaner results may be obtained when considering infinite sequences. This is apparent from the extra care and combinatorial arguments employed to handle finite words and non-shift-invariant measures. We leave this study of infinite sequences to a later work.

Another set of open questions raised by this work is the study of various complexity properties associated with encoders for SCSs. In the two encoders presented here, we briefly mentioned number of states and edges, as well as anticipation, as important parameters. These are crucial for practical applications. A more in-depth study of these and other parameters, as well as associated bounds and trade-offs between them, will by the subject of future work.

Acknowledgments

The authors would like to thank the associate editor and the anonymous reviewers, whose comments helped improve the presentation of the paper.

References

[1] J.-R. Chazottes, J.-M. Gambaudo, M. Hochman, and E. Ugalde, "On the finite-dimensional marginals of shift-invariant measures," *Ergodic Theory Dyn. Syst.*, vol. 32, no. 5, pp. 1485–1500, 2012.

[2] Y. M. Chee, C. Johan, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Efficient encoding/decoding of capacity-achieving constant-composition ICI-free codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 205–209.

[3] Y. M. Chee, C. Johan, H. M. Kiah, S. Ling, T. T. Nguyen, and V. K. Vu, "Rates of constant-composition codes that mitigate intercell interference," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 200–204.

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY, USA: Springer, 1998.

[5] O. Elishco, T. Meyerovitch, and M. Schwartz, "Semiconstrained systems," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 811–824, Apr. 2016.

[6] T. Etzion, "Cascading methods for runlength-limited arrays," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 319–324, Jan. 1997.

[7] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Eindhoven, The Netherlands: Shannon Foundation Publishers, 2004.

[8] R. Karabed, D. L. Neuhoff, and A. Khayrallah, "The capacity of costly noiseless channels," Almaden Res. Center, San Jose, CA, USA, Tech. Rep. RJ 6040 (59639), Jan. 1988.

[9] S. Kayser and P. H. Siegel, "Constructions for constant-weight ICI-free codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 1431–1435.

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

[11] V. Y. Krachkovsky, R. Karabed, S. Yang, and B. A. Wilson, "On modulation coding for channels with cost constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 421–425.

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

[13] D. Lind and B. H. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

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

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

[16] 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. Commun.*, vol. 55, no. 2, pp. 237–241, Feb. 2007.

[17] J. Shallit, *A Second Course in Formal Languages and Automata Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[18] H. J. Shyr, "Characterizations of right dense languages," *Semigroup Forum*, vol. 33, no. 1, pp. 23–30, 1986.

[19] J. B. Soriaga and P. H. Siegel, "On the design of finite-state shaping encoders for partial-response channels," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2006, pp. 1–5.

Ohad Elishco (S'12) received the B.Sc. in mathematics and his B.Sc. in electrical engineering in 2012 from Ben-Gurion University of the Negev, Israel; the M.Sc. degree in electrical engineering in 2013 from Ben-Gurion University of the Negev. In 2013 he started his Ph.D in electrical engineering also in Ben-Gurion University. His research interests are constrained coding and dynamical systems.

Tom Meyerovitch received the B.Sc. in mathematics and computer science from from Tel-Aviv University, Israel, in 1999; the M.Sc. and Ph.D. degrees in mathematics in 2004 and 2010, respectively, also from Tel-Aviv University. In 2010-2012 he held a post doctoral position at the University of British Columbia, and the Pacific Institute for Mathematical studies in Vancouver Canada. Since 2012, he has been on the faculty of the Department of Mathematics at Ben-Gurion University, Beer-Sheva, Israel. He studies various mathematical aspects of dynamical systems, in particular symbolic dynamics.

Moshe Schwartz (M'03–SM'10) is an associate professor at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz received the B.A. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Technion – Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004 respectively, all from the Computer Science Department. He was a Fulbright post-doctoral researcher in the Department of Electrical and Computer Engineering, University of California San Diego, and a post-doctoral researcher in the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012–2014, he was a visiting scientist at the Massachusetts Institute of Technology (MIT).

Prof. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage.