

On the Capacity of Constrained Systems

G. Böcherer, R. Mathar

*Institute for Theoretical Information Technology
RWTH Aachen University
52056 Aachen, Germany
Email: {boecherer,mathar}@ti.rwth-aachen.de

V.C. da Rocha Jr., C. Pimentel

Communications Research Group - CODEC
Department of Electronics and Systems, P.O. Box 7800
Federal University of Pernambuco
50711-970 Recife PE, Brazil
E-mail: {vcr,cecilio}@ufpe.br

Abstract—In the first chapter of Shannon’s *A Mathematical Theory of Communication*, it is shown that the maximum entropy rate of an input process of a constrained system is limited by the combinatorial capacity of the system. Shannon considers systems where the constraints define regular languages and uses results from matrix theory in his derivations. In this work, the regularity constraint is dropped. Using generating functions, it is shown that the maximum entropy rate of an input process is upper-bounded by the combinatorial capacity in general. The presented results also allow for a new approach to systems with regular constraints. As an example, the results are applied to binary sequences that fulfill the (j, k) run-length constraint and by using the proposed framework, a simple formula for the combinatorial capacity is given and a maxentropic input process is defined.

I. INTRODUCTION

This work is motivated by the recent interest in the information-theoretic limits of systems with constraints that do not form a regular language. One example is the consideration of context-free languages with an application to genetic sequence modelling [1], another example is the investigation of asynchronous channels [2].

A constrained system allows the noiseless transmission of input sequences of weighted symbols that fulfill certain constraints on the symbol constellations. A natural question is how to efficiently encode a random source such that it becomes a valid input for a constrained system [3]. Furthermore, it is of interest to determine the ultimate performance of such an encoder, which is closely related to the entropy rate of random processes that generate strings that fulfill the constraints. The maximum entropy rate of all such processes is equal to the combinatorial capacity of the considered system in the case that the constraints form a regular language. This was originally shown in [4]. In [5], the authors show this property for a slightly generalized setup, since they allow non-integer valued symbol weights, as long as the set of weights is not too dense. We will define what “not too dense” means in Section III. Recently, the authors of [2] showed that combinatorial capacity and maximum entropy rate are equal for a specific class of constrained systems, which they call the asynchronous channel. It is worthy to note that for

asynchronous channels, the “not too dense” assumption is not necessarily fulfilled.

In this work, we consider constrained systems with not necessarily regular constraints; we allow symbol weights taking arbitrary positive real values, and motivated by [2], we allow the set of symbol weights to be too dense. For this general class of constrained systems, we show how to represent such systems by generating functions. We give a new definition of combinatorial capacity that coincides with the original definition [4] when the weight set is not too dense. By invoking known results, we show that the combinatorial capacity of such systems is equal to the abscissa of convergence of the corresponding generating function. Finally, we define input processes of constrained systems and show that the maximum entropy rate of input processes is upper-bounded by the abscissa of convergence of the generating function. This is our main result: independent from if the “not too dense” property is fulfilled or not, and for whatever constraints, the combinatorial capacity is equal to the abscissa of convergence of the generating function, and the entropy rate is upper-bounded by the abscissa of convergence of the generating function. By a detailed discussion of the (j, k) run-length constraint, we illustrate our ideas and we show that our framework, besides being more general, also allows for a new approach to investigate regular systems. Namely, we derive a simple formula for the combinatorial capacity of the (j, k) constraint and we define an input process whose entropy rate is equal to the combinatorial capacity.

The remainder of this paper is organized as follows. In Section II, we define constrained systems and generating functions. In Section III, we show how to calculate the combinatorial capacity. We then show in Section IV how the combinatorial capacity relates to entropy rates and finally, in Section V, we show that the entropy rate of input processes is upper-bounded by the combinatorial capacity.

II. CONSTRAINED SYSTEMS

In this section, we define the class of constrained systems that we will investigate in this work and we show how to represent them by generating functions.

Definition 1. A *constrained system* $A = (\mathcal{A}, w)$ consists of a countable set \mathcal{A} of strings accepted by the system and an

This work has been supported by the UMIC Research Centre, RWTH Aachen University. This work has also been supported by the Brazilian National Council for Scientific and Technological Development (CNPq).

A system $S_{(j,k)} = (\mathcal{A}, w)$ accepting binary strings that fulfill the (j, k) constraint can be defined as follows using regular expressions [6]:

$$\begin{aligned} \mathcal{A} = & (1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}}) [(0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}}) (1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}})]^* (\varepsilon \cup 0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}}) \\ & \cup (0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}}) [(1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}}) (0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}})]^* (\varepsilon \cup 1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}}) \end{aligned} \quad (\text{E.1})$$

$$w(0) = w(1) = 1. \quad (\text{E.2})$$

The symbol $*$ denotes the Kleene star, \cup denotes the or-operation, ε denotes the empty-string with $w(\varepsilon) = 0$, 11 denotes “1 concatenated with 1”. Note that concatenation is non-commutative: $1(1 \cup 0) = 11 \cup 10 \neq 11 \cup 01$.

Example 1. Definition of $S_{(j,k)}$.

From (E.1) and (E.2) in Example 1, the generating function of $S_{(j,k)}$ can directly be derived as

$$\begin{aligned} G_{(j,k)}(s) = & (e^{-s} + \dots + e^{-js}) \sum_{n=0}^{\infty} [(e^{-s} + \dots + e^{-ks})(e^{-s} + \dots + e^{-js})]^n (1 + e^{-s} + \dots + e^{-ks}) \\ & + (e^{-s} + \dots + e^{-ks}) \sum_{n=0}^{\infty} [(e^{-s} + \dots + e^{-js})(e^{-s} + \dots + e^{-ks})]^n (1 + e^{-s} + \dots + e^{-js}). \end{aligned} \quad (\text{E.3})$$

A discussion of how to derive generating functions from regular expressions can be found in [7].

Example 2. Generating function of $S_{(j,k)}$.

associated weight function $w: \mathcal{A} \rightarrow \mathbb{R}_{>0}$ ($\mathbb{R}_{>0}$ denotes the positive real numbers) with the following property: if $a, b \in \mathcal{A}$ and $ab \in \mathcal{A}$ then $w(ab) = w(a) + w(b)$.

The weight of a symbol can have different practical meaning. In the context of magnetic recording systems, “weight” will probably refer to “tape-length”; other meanings like “time” or “energy” are possible, depending on the modelled system. For an illustration of our definition, we define in Example 1 the system $S_{(j,k)}$, which accepts binary sequences that consist of at most j consecutive 1s and at most k consecutive 0s. This constrained is called the (j, k) run-length constraint. We return to $S_{(j,k)}$ at several points in our paper.

A. Generating Functions

To analyze the asymptotic behavior of constrained systems, we first represent the set \mathcal{A} of allowed strings together with the weight function w by a generating function. We then interpret the generating function as a function on the complex plane and investigate its convergence behavior. This approach, mostly referred to as *analytic combinatorics*, is discussed in detail in [8]. We consider a more general case since we do not restrict the range of the weight function w to the natural numbers, but allow for the set of positive real numbers $\mathbb{R}_{>0}$. Therefore, we use general Dirichlet series [9] instead of power series as generating functions.

Definition 2. Let $A = (\mathcal{A}, w)$ represent a constrained system. We define the *generating function* of A by

$$G_A(s) = \sum_{a \in \mathcal{A}} e^{-w(a)s}, \quad s \in \mathbb{C} \quad (1)$$

where \mathbb{C} denotes the set of complex numbers.

Let Ω denote the set of distinct string weights of elements in \mathcal{A} . We order and index the set Ω such that $\Omega = \{\nu_k\}_{k=1}^{\infty}$ with $\nu_1 < \nu_2 < \dots$. For every $\nu_k \in \Omega$, $N(\nu_k)$ denotes the number of distinct strings of weight ν_k in \mathcal{A} . We can now write the generating function as

$$G_A(s) = \sum_{k=1}^{\infty} N(\nu_k) e^{-\nu_k s}. \quad (2)$$

Since the coefficients $N(\nu_k)$ result from an enumeration, they are all non-negative. In Example 2, we show how to represent $S_{(j,k)}$ by a generating function.

III. COMBINATORIAL CAPACITY

In previous works that consider non-integer valued weights [5], [10], the authors restrict themselves to constrained systems where the ordered set of string weights $\Omega = \{\nu_k\}_{k=1}^{\infty}$ is *not too dense*, that is, there exists some constant $L \geq 0$ and some constant $K \geq 0$ such that for any integer $n \geq 0$

$$\max_{\nu_k < n} k \leq Ln^K. \quad (3)$$

Under the “not too dense” assumption, it is meaningful to follow Shannon’s original definition and to identify the combinatorial capacity with C_0 given by

$$C_0 = \limsup_{k \rightarrow \infty} \frac{\ln N(\nu_k)}{\nu_k} \quad (4)$$

as was done for instance in [5] and [10]. Here and hereafter, \ln denotes the natural logarithm. Throughout the paper, for

a sequence $\{s_l\}_{l=1}^{\infty}$, $s = \limsup_{l \rightarrow \infty} s_l$ is equivalent to the following: for any $\epsilon > 0$, it holds that

$$s_l \leq s + \epsilon \quad \text{almost everywhere (a.e.)} \quad (5)$$

and

$$s_l \geq s - \epsilon \quad \text{infinitely often (i.o.)} \quad (6)$$

with respect to $l \in \mathbb{N}$ (\mathbb{N} denotes the set of natural numbers starting with one).

If Ω is too dense, the number of possible string weights in the interval $[n, n + 1]$ increases faster than polynomial with n , in which case identifying combinatorial capacity with C_0 may become inappropriate. See [10] for an example and see [5] for a detailed discussion.

We now give a definition of combinatorial capacity that is meaningful also when the ordered set of string weights Ω is too dense.

Definition 3. We define the *combinatorial capacity* C as

$$C = \limsup_{k \rightarrow \infty} \frac{\ln \left[\sum_{l=1}^k N(\nu_l) \right]}{\nu_k}. \quad (7)$$

The motivation for our generalized definition is twofold. First, our results on entropy rates for constrained systems, which we will present in the remaining sections, do not depend on the “not too dense” property. Second, recent work [2] has shown that there are constrained systems of practical interest that may not necessarily have the “not too dense” property. The following theorem shows that Definition 3 of combinatorial capacity is consistent with the conventional definition (4), namely, if Ω is not too dense, then our definition of combinatorial capacity coincides with (4), i.e., $C = C_0$.

Theorem 1. Let $A = (\mathcal{A}, w)$ be a constrained system with the set of distinct string weights Ω and generating function $G_A(s)$. Denote the abscissa of convergence of G_A by Q . The following holds:

(i) The combinatorial capacity C is equal to Q , i.e.,

$$\limsup_{k \rightarrow \infty} \frac{\ln \left[\sum_{l=1}^k N(\nu_l) \right]}{\nu_k} = Q. \quad (8)$$

(ii) If the set of distinct weights Ω is not too dense, then $C_0 = C$, i.e.,

$$\limsup_{k \rightarrow \infty} \frac{\ln N(\nu_k)}{\nu_k} = \limsup_{k \rightarrow \infty} \frac{\ln \left[\sum_{l=1}^k N(\nu_l) \right]}{\nu_k}. \quad (9)$$

Proof: We start by proving statement (i). All coefficients $N(\nu_k)$ are non-negative since they result from an enumeration. Therefore, for all $k \in \mathbb{N}$, $N(\nu_k) = |N(\nu_k)|$. With this observation, (i) follows directly from [9, Theorem 7]. To proof statement (ii), we assume that Ω is not too dense. In this case $C_0 = Q$, which was shown in [10, Lemma 1]. Statement (ii) now follows from statement (i). Alternatively, (ii) can be shown by combinatorial arguments, see [7, Appendix B.3] ■

By Theorem 1, the combinatorial capacity of $S_{(j,k)}$ is given by the abscissa of convergence of its generating function $G_{(j,k)}$. From (E.3) in Example 1, we see that the abscissa of convergence of $G_{(j,k)}$ is given by the largest positive real solution of

$$(e^{-s} + \dots + e^{-js})(e^{-s} + \dots + e^{-ks}) = 1. \quad (\text{E.4})$$

This formula coincides with the formula given in [11, Theorem 2] and it can also be derived by applying the techniques introduced in [12].

Example 3. Combinatorial capacity of $S_{(j,k)}$.

Returning to our example, the set of symbol weights of $S_{(j,k)}$ is not too dense, because the underlying alphabet $\{0, 1\}$ of \mathcal{A} is finite. See [5, Appendix A] for a more detailed discussion of this argument. Using Theorem 1, we derive in Example 3 a simple formula for the combinatorial capacity of $S_{(j,k)}$.

IV. THE RELATION BETWEEN COMBINATORIAL CAPACITY AND ENTROPY RATES

After having defined the combinatorial capacity of constrained systems in the last section, we now want to consider random processes that generate strings that fulfill the constraints of the considered system. We then want to know how the maximum entropy rate of such a process relates to the combinatorial capacity. Ultimately, we have a process in mind that generates at each time instant a substring, which is then appended to the string that has been generated so far, such that at each time instant, the generated string fulfills the constraints of the system. In magnetic recording, such a process would generate a substring, write it to the tape, generate another substring, write it to the tape, and so forth, without ever rewinding the tape. The difficulty of analyzing such a process is the following: fix two time instants l and l' , $l < l'$. The probability that the process writes a specific string to the tape until time instant l' depends in general on the probabilities of the strings that it can write until time instant l . This dependency can become arbitrarily complicated depending on the constraints of the system. Because of these interdependencies, it is difficult to bound the entropy rate of such a process. We solve these interdependencies by decoupling the time instants l and l' : each time the recording system wants to write to the tape, it first rewinds the tape completely and then overwrites everything that has been written before. We call such a system an input source (in contrast to an input process) of a constrained system. In this section, we show through a series of results that the entropy rate of input sources is upper-bounded by the combinatorial capacity and we postpone input processes until Section V.

A. Input Sources for Constrained Systems

Definition 4. Let $A = (\mathcal{A}, w)$ denote a constrained system. Denote by $X = \{X_l\}_{l=1}^{\infty}$ a sequence of random variables and

The sequence of random variables $\{X_l\}_{l=1}^{\infty}$ with support of X_l given by

$$\mathcal{X}_l = [(0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}})(1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}})]^l \quad (\text{E.5})$$

is an input source of A : first, $\bigcup_{l=1}^{\infty} \mathcal{X}_l \subseteq \mathcal{A}$ and $\mathcal{X}_l \cap \mathcal{X}_k = \emptyset$ whenever $l \neq k$ and second, $\mathcal{X}_l \neq \emptyset$ for each $l \in \mathbb{N}$, which shows that both condition (i) and condition (ii) of Definition 4 are fulfilled.

Example 4. An input source for $S_{(j,k)}$.

denote by \mathcal{X}_l the support of X_l . We say that X is an *input source* of A if and only if

- (i) $\bigcup_{l=1}^{\infty} \mathcal{X}_l \subseteq \mathcal{A}$ and $\mathcal{X}_l \cap \mathcal{X}_k = \emptyset$, if $l \neq k$.
- (ii) For each $l \in \mathbb{N}$, $\mathcal{X}_l \neq \emptyset$.

We define in Example 4 an input source for $S_{(j,k)}$. Note that the given example is not the only possible input source of $S_{(j,k)}$.

We denote the *probability mass function* (PMF) of X_l by

$$p_{X_l}(x) = \mathbb{P}[X_l = x], \quad x \in \mathcal{X}_l. \quad (10)$$

Definition 5. We define the *entropy rate* \bar{H} of an input source X by

$$\bar{H}(X) = \limsup_{l \rightarrow \infty} \frac{H(X_l)}{\mathbb{E}[w(X_l)]} \quad (11)$$

where $\mathbb{E}[w(X_l)]$ denotes the average weight of all $x \in \mathcal{X}_l$ with respect to the PMF p_{X_l} and where $H(X_l)$ denotes the entropy of X_l in nats.

We can upper-bound the entropy rate $\bar{H}(X)$ by maximizing each term of the sequence on the right-hand side of (11) separately. To do so, we need the following lemma:

Lemma 1. Denote by p_Z the PMF of some random variable Z with countable support \mathcal{Z} and an associated positive weight function w . The maximum entropy per average weight

$$R_Z = \max_{p_Z} \frac{H(Z)}{\mathbb{E}[w(Z)]} \quad (12)$$

is given by the greatest positive real solution of the equation

$$\sum_{z \in \mathcal{Z}} e^{-w(z)s} = 1. \quad (13)$$

In addition, the PMF of Z that achieves this rate is uniquely given by

$$q_Z(z) = e^{-w(z)R_Z}, \quad z \in \mathcal{Z}. \quad (14)$$

Proof: These two properties of R_Z were derived by using Lagrange Multipliers in [13] and they were independently derived in [14] by using the bound $\ln z \leq z - 1$. We offer an alternative proof by applying the information inequality [15], which states for the Kullback Leibler Distance $D(\cdot \parallel \cdot)$ of two PMFs p and q that

$$D(p \parallel q) \geq 0 \quad (15)$$

with equality if and only if $p = q$. We thus have

$$0 \geq -D(p_Z \parallel q_Z) \quad (16)$$

$$= \sum_{z \in \mathcal{Z}} p_Z(z) \ln \frac{q_Z(z)}{p_Z(z)} \quad (17)$$

$$= H(Z) - R_Z \mathbb{E}[w(Z)] \quad (18)$$

which implies

$$\frac{H(Z)}{\mathbb{E}[w(Z)]} \leq R_Z \quad (19)$$

with equality if and only if $p_Z = q_Z$. ■

Lemma 2. Let X denote an input source of some constrained system. Let the rate bound R_X be defined as

$$R_X = \limsup_{l \rightarrow \infty} R_{X_l} \quad (20)$$

where each R_{X_l} is chosen according to Lemma 1. The entropy rate $\bar{H}(X)$ of X is then upper-bounded by R_X .

Proof: We have

$$\bar{H}(X) = \limsup_{l \rightarrow \infty} \frac{H(X_l)}{\mathbb{E}[w(X_l)]} \quad (21)$$

$$\leq \limsup_{l \rightarrow \infty} \max_{p_{X_l}} \frac{H(X_l)}{\mathbb{E}[w(X_l)]} \quad (22)$$

$$= \limsup_{l \rightarrow \infty} R_{X_l} \quad (23)$$

$$= R_X. \quad (24)$$

Lemma 3. Let $A = (\mathcal{A}, w)$ represent a constrained system and let X denote an input source of A . The rate bound R_X of X is then upper-bounded by the abscissa of convergence Q of G_A , i.e.,

$$R_X \leq Q. \quad (25)$$

Proof: To prove the lemma, we show that the generating function $G_A(s)$ diverges whenever $\text{Re}(s) < R_X$, for any input source X .

The definition of the rate bound R_X in (20) implies in particular that for any $\epsilon > 0$

$$R_{X_l} \geq R_X - \epsilon \quad \text{i.o.} \quad (26)$$

According to (13), R_{X_l} is given by the greatest positive real solution of

$$\sum_{x \in \mathcal{X}_l} e^{-w(x)s} = 1 \quad (27)$$

which implies further

$$\sum_{x \in \mathcal{X}_l} e^{-w(x)[R_X - \epsilon]} \geq \sum_{x \in \mathcal{X}_l} e^{-w(x)R_{X_l}} = 1 \quad \text{i.o.} \quad (28)$$

Because of $\bigcup_{l=1}^{\infty} \mathcal{X}_l \subseteq \mathcal{A}$ according to Definition 4, we can bound the generating function by

$$G_A(s) = \sum_{a \in \mathcal{A}} e^{-w(a)s} \geq \lim_{n \rightarrow \infty} \sum_{l=1}^n \sum_{x \in \mathcal{X}_l} e^{-w(x)s}. \quad (29)$$

Because of (28), we have

$$\sum_{l=1}^n \sum_{x \in \mathcal{X}_l} e^{-w(x)[R_X - \epsilon]} \xrightarrow{n \rightarrow \infty} \infty \quad (30)$$

and we conclude that for any $\epsilon > 0$, G_A diverges in $s = R_X - \epsilon$. Thus, by [9, Theorem 3], G_A diverges for all $s \in \mathbb{C}$ with $\text{Re}(s) < R_X$. Since by definition of Q , G_A converges for $\text{Re}(s) > Q$, it must hold that $R_X \leq Q$. ■

V. MAXENTROPIC INPUT PROCESSES

We now come to the main concern of this work: we want to define input processes for constrained systems and we want to investigate how the entropy rate of an input process is related to the combinatorial capacity. Loosely speaking, we want to define a random process that generates at each time instant a substring, which is then appended to the string that has been generated so far. At each time instant, the complete string should fulfill the constraints of the considered system. The notion of an input process differs fundamentally from what we defined as an input *source*: an input source generates a complete new string at each time instant. Before we give our definition of input processes, we motivate our definition by the following example. Consider a constrained system S_{bin} that accepts any binary sequence, and assume $w(1) = w(0) = 1$. The combinatorial capacity is $C = \ln(2) \approx 0.6932$. Denote by $V = \{V_l\}_{l=1}^{\infty}$ the random process where the V_l take values in $\{0, 1, 01\}$ and are independent, identically distributed (IID) according to the PMF p_V , which we define as follows:

$$p_V(0) = p_V(1) = e^{-R}, \quad p_V(01) = e^{-2R} \quad (31)$$

where R is given by the largest positive real solution of

$$2e^{-s} + e^{-2s} = 1. \quad (32)$$

Obviously, V generates binary strings that are accepted by S_{bin} . We are interested in the entropy rate of V and calculate

$$\frac{H(V)}{E[w(V)]} = R \quad (33)$$

$$\approx 0.8814 \quad (34)$$

$$> C \quad (35)$$

where equality in (33) follows from Lemma 1. Surprisingly, the entropy rate of V seems to exceed the combinatorial capacity of S_{bin} . The reason for this is that we implicitly assume in our attempt to calculate the entropy rate of V that, for example, the realizations $v_1 = 01$ and $(v_1, v_2) = (0, 1)$ are distinguishable, however, they are not: both result in the string 01 , so we are counting this string twice. To avoid this pitfall, we define input processes as follows.

Definition 6. The random process $\{Y_l\}_{l=1}^{\infty}$, $Y_l \in \mathcal{Y}$ is an *input process* of the constrained system $A = (\mathcal{A}, w)$ if the sequence of random variables $X_l = \text{cat}(Y_1, \dots, Y_l)$ with the supports truncated to

$$\mathcal{X}_l = \{\text{cat}(y_1, \dots, y_l) | (y_1, \dots, y_l) \in \mathcal{Y}^l, p_Y(y_1, \dots, y_l) > 0\} \quad (36)$$

is an input source of A . The operator cat denotes concatenation: $\text{cat}(a, b) = ab$.

We refer to the assignment (36) in the following by *truncated support*.

The process V as we defined it earlier is not an input process of S_{bin} : define $X_l = \text{cat}(V_1, \dots, V_l)$, $l = 1, 2, \dots$. The random variables X_1 and X_2 have the following truncated supports:

$$\mathcal{X}_1 = \{0, 1, 01\} \quad (37)$$

$$\mathcal{X}_2 = \{00, 01, 001, 10, 11, 101, 010, 011, 0101\}. \quad (38)$$

As we can see, $\mathcal{X}_1 \cap \mathcal{X}_2 = 01 \neq \emptyset$, so X is not an input source of S_{bin} and thus, by definition, V is not an input process of S_{bin} . By changing the PMF of the V_l to

$$p_V(0) = p_V(1) = \frac{1}{2}, \quad p_V(01) = 0 \quad (39)$$

each random variable X_l has the truncated support

$$\mathcal{X}_l = (0 \cup 1)^l. \quad (40)$$

Thus, with the probability assignment (39), $\mathcal{X}_l \cap \mathcal{X}_k = \emptyset$ whenever $k \neq l$, and consequently, X is an input source and Y is an input process.

Now that we have defined input processes in a way that does not allow for “counting twice”, we can give the following

Definition 7. The *entropy rate* of an input process Y is defined as

$$\bar{H}(Y) = \limsup_{l \rightarrow \infty} \frac{H(Y_1, \dots, Y_l)}{E[w(Y_1) + \dots + w(Y_l)]}. \quad (41)$$

The entropy rate of an input process of a constrained system relates to the combinatorial capacity as follows.

Theorem 2. Let $A = (\mathcal{A}, w)$ represent a constrained system. The entropy rate of an input process Y of A is upper-bounded by the abscissa of convergence Q of G_A , and in particular, it is upper bounded by the combinatorial capacity C of A .

Proof: Since Y is an input process of A , by definition, the sequence X of random variables $X_l = (Y_1, \dots, Y_l)$ with truncated supports is an input source of A . We thus have

$$\bar{H}(Y) = \limsup_{l \rightarrow \infty} \frac{H(Y_1, \dots, Y_l)}{E[w(\text{cat}(Y_1, \dots, Y_l))]} \quad (42)$$

$$= \limsup_{l \rightarrow \infty} \frac{H(X_l)}{E[w(X_l)]} \quad (43)$$

$$\leq R_X \quad (44)$$

$$\leq Q \quad (45)$$

$$= C \quad (46)$$

where the equality in (43) follows from the definition of X , the inequality in (44) follows from Lemma 2, the inequality in (45) follows from Lemma 3, and the equality in (46) follows from Theorem 1. ■

We call an input process Y^* of a constrained system A *maxentropic* if for any input process Y of A we have $\bar{H}(Y^*) \geq$

Let $Y = \{Y_l\}_{l=1}^{\infty}$ denote a random process with realizations $(y_1, \dots, y_l) \in \mathcal{Y}^l$. Let \mathcal{Y} be given by

$$\mathcal{Y} = (0 \cup \dots \cup \underbrace{0 \dots 0}_{k \text{ times}})(1 \cup \dots \cup \underbrace{1 \dots 1}_{j \text{ times}}). \quad (\text{E.6})$$

From Example 4, we know that the sequence of random variables $\{X_l\}_{l=1}^{\infty}$ with $X_l = \text{cat}(Y_1, \dots, Y_l)$ is an input source of $S_{(j,k)}$; therefore, Y is an input process of $S_{(j,k)}$. Let $\{Y_l\}_{l=1}^{\infty}$ be independent, identically distributed (IID) according to the PMF

$$p_Y(y) = e^{-w(y)R}, \quad y \in \mathcal{Y} \quad (\text{E.7})$$

where R is given by the largest positive real solution of

$$\sum_{y \in \mathcal{Y}} e^{-w(y)s} = (e^{-s} + \dots + e^{-js})(e^{-s} + \dots + e^{-ks}) = 1. \quad (\text{E.8})$$

Comparing (E.8) with (E.4) in Example 3, we see that $R = C$, i.e., R is equal to the combinatorial capacity C of $S_{(j,k)}$. The entropy rate of Y is

$$\bar{H}(Y) = \limsup_{l \rightarrow \infty} \frac{H(Y_1, \dots, Y_l)}{E[w(\text{cat}(Y_1, \dots, Y_l))]} \quad (\text{E.9})$$

$$= \limsup_{l \rightarrow \infty} \frac{l H(Y)}{l E[w(Y)]} \quad (\text{E.10})$$

$$= R = C \quad (\text{E.11})$$

where the second equality follows from the independence bound on entropy [15] together with the fact that the Y_l are IID, and the linearity of w . Since, from Theorem 2, $\bar{H}(Y) \leq C$, we conclude that Y is a maxentropic input process of $S_{(j,k)}$.

Example 5. A maxentropic input process of $S_{(j,k)}$

$\bar{H}(Y)$. Because of Theorem 2, $\bar{H}(Y^*) = C$, i.e., the entropy rate of Y^* is equal to the combinatorial capacity of A , is a sufficient condition for Y^* to be maxentropic. It is important to note that Theorem 2 does not claim that $\bar{H}(Y^*) = C$ for any constrained system A . For a large class of constrained systems, however, $\bar{H}(Y^*) = C$. For this class, Theorem 2 is quite useful: assume that we want to show that system A belongs to this class. Without Theorem 2, we have to maximize the entropy rate over all input processes of A . Once we have determined the maximum entropy rate, we compare it to the combinatorial capacity and find that both are equal. An example of this approach can be found in the proof of [5, Theorem 5.1]. With Theorem 2, we can do something different: we look for an input process whose entropy rate is equal to the combinatorial capacity. Once we have found such an input process, we invoke

Theorem 2 and are done. We illustrate this new approach in Example 5.

VI. CONCLUSIONS

In this work, we showed for a general class of constrained systems (including those with non-regular constraints and dropping the “not too dense” assumption for the weight set) that the maximum entropy rate of input processes is upper-bounded by the combinatorial capacity of the considered system. This general result allows for a new approach to show that maximum entropy rate and combinatorial capacity are *equal*: with our result, it is enough to find “some” input process whose entropy rate is equal to the combinatorial capacity of the considered system. Equality of *maximum* entropy rate and combinatorial capacity then follows from our result. In contrast to previous works (except for some works that consider specific classes of constraint systems), we do not use any result from matrix theory in our derivations. Our framework, which is based on generating functions, therefore allows, besides being more general, for a new approach to investigate regular systems. We illustrated this by applying our results to the (j, k) constraint.

REFERENCES

- [1] O. Milenkovic, “Constrained coding for context-free languages with applications to genetic sequence modelling,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2007, pp. 1686–1690.
- [2] R. Yeung, N. Cai, S.-W. Ho, and A. Wagner, “Reliable communication in the absence of a common clock,” *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 700–712, 2009.
- [3] B. H. Marcus, R. M. Roth, and P. H. Siegel, “An introduction to coding for constrained systems,” Oct. 2001.
- [4] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.
- [5] A. Khandekar, R. McEliece, and E. Rodemich, “The discrete noiseless channel revisited,” in *Coding, Communications, and Broadcasting*. Research Studies Press Ltd., 2000, pp. 115–137.
- [6] M. Sipser, *Introduction to the Theory of Computation*, 2nd ed. Thomson Course Technology, 2006.
- [7] G. Bocherer, “Analytic asymptotics of discrete noiseless channels,” Master’s thesis, ETH Zurich, 2007. [Online]. Available: <http://www.georg-boecherer.de/repository/analyticAsymptotics.pdf>
- [8] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge University Press, 2008.
- [9] G. H. Hardy and M. Riesz, *The General Theory of Dirichlet’s Series*. Cambridge: at the University Press, 1915.
- [10] G. Bocherer, V. C. da Rocha Jr., and C. Pimentel, “Capacity of general discrete noiseless channels,” in *Proc. Int. Symp. Commun. Applicat. (ISCTA)*, 2007. [Online]. Available: <http://arxiv.org/abs/0802.2451>
- [11] T. Mittelholzer, “Enumerative maximum-transition-run codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2009, pp. 1549–1553.
- [12] C. Pimentel and B. Uchoa-Filho, “A combinatorial approach to finding the capacity of the discrete noiseless channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 2024–2028, 2003.
- [13] R. S. Marcus, “Discrete noiseless coding,” Master’s thesis, Massachusetts Institute of Technology, 1957.
- [14] R. M. Krause, “Channels which transmit letters of unequal duration,” *Inf. Contr.*, vol. 5, pp. 3–24, 1962.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc, 2006.